U.S. Department of
Transportation

**Federal Railroad
Administration**

Quantification of the Sensitivity of Two Prevalent Track Inspection Systems

Office of Research,
Development,
and Technology
Washington, DC 20590

DOT/FRA/ORD-17/03

Final Report
May 2017

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2017	Technical Report

4. TITLE AND SUBTITLE
Quantification of the Sensitivity of Two Prevalent Track Inspection Systems

5. FUNDING NUMBERS

6. AUTHOR(S)
Leith Al-Nazer[1], Thomas Raslear[1], Robert Wilson[1], John Kidd[2]

[1]Federal Railroad Administration; [2]QinetiQ North America

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Data Collection: QinetiQ North America
350 2nd Avenue
Waltham, MA 02451
Data Analysis: U.S. Department of Transportation
Federal Railroad Administration
Office of Research, Development, and Technology
1200 New Jersey Avenue, SE
Washington, DC 20590

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Railroad Administration
Office of Railroad Policy and Development
Office of Research and Development
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DOT/FRA/ORD-17/03

11. SUPPLEMENTARY NOTES
COR: Leith Al-Nazer

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the FRA Web site at http://www.fra.dot.gov.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
This report uses signal detection theory to quantify the sensitivity of two common track inspection systems (manual visual inspection via a hi-rail vehicle and automated track geometry inspection). Quantifying the sensitivity of such systems can provide baseline input data for risk-based, probabilistic track inspection programs, which cannot determine optimal track inspection methodologies without accurate probabilities of detection for various track inspection methods. The results show that there are no overall differences in sensitivity (ability to detect deteriorated or anomalous track conditions) between visual and automated inspection methods. However, there is a strong bias with both systems to indicate "no defect," which results in a high rate of misses of deteriorated and anomalous track conditions.

14. SUBJECT TERMS
Track inspection, inspection systems, inspection methodologies, optimal inspection, risk-based inspection, signal detection theory

15. NUMBER OF PAGES
69

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in) = 2.5 centimeters (cm)
1 foot (ft) = 30 centimeters (cm)
1 yard (yd) = 0.9 meter (m)
1 mile (mi) = 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2) = 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2) = 0.09 square meter (m^2)
1 square yard (sq yd, yd^2) = 0.8 square meter (m^2)
1 square mile (sq mi, mi^2) = 2.6 square kilometers (km^2)
1 acre = 0.4 hectare (he) = 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz) = 28 grams (gm)
1 pound (lb) = 0.45 kilogram (kg)
1 short ton = 2,000 pounds (lb) = 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp) = 5 milliliters (ml)
1 tablespoon (tbsp) = 15 milliliters (ml)
1 fluid ounce (fl oz) = 30 milliliters (ml)
1 cup (c) = 0.24 liter (l)
1 pint (pt) = 0.47 liter (l)
1 quart (qt) = 0.96 liter (l)
1 gallon (gal) = 3.8 liters (l)
1 cubic foot (cu ft, ft^3) = 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3) = 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

[(x-32)(5/9)] °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm) = 0.04 inch (in)
1 centimeter (cm) = 0.4 inch (in)
1 meter (m) = 3.3 feet (ft)
1 meter (m) = 1.1 yards (yd)
1 kilometer (km) = 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2) = 0.16 square inch (sq in, in^2)
1 square meter (m^2) = 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2) = 0.4 square mile (sq mi, mi^2)
10,000 square meters (m^2) = 1 hectare (ha) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gm) = 0.036 ounce (oz)
1 kilogram (kg) = 2.2 pounds (lb)
1 tonne (t) = 1,000 kilograms (kg)
= 1.1 short tons

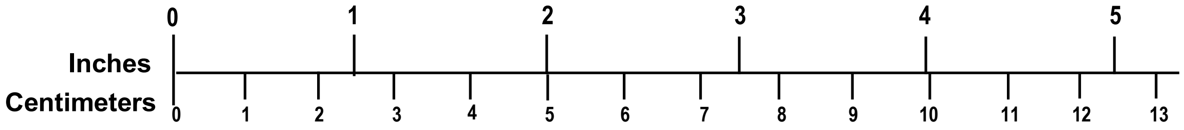
VOLUME (APPROXIMATE)

1 milliliter (ml) = 0.03 fluid ounce (fl oz)
1 liter (l) = 2.1 pints (pt)
1 liter (l) = 1.06 quarts (qt)
1 liter (l) = 0.26 gallon (gal)
1 cubic meter (m^3) = 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3) = 1.3 cubic yards (cu yd, yd^3)

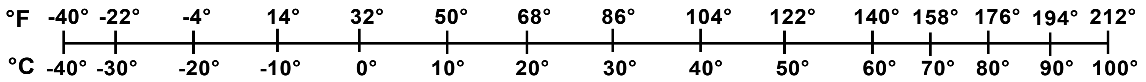
TEMPERATURE (EXACT)

[(9/5) y + 32] °C = x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSIO



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

Acknowledgements

The authors would like to thank all the track inspectors in FRA's Office of Railroad Safety who participated in the data collection field tests.

In addition, the authors would like to thank officials of the American Federation of Government Employees Local 2814 for their support of this study.

Last but not least, the authors would like to thank officials and track personnel at the three participating railroads for supporting the study, providing track time, and being patient when last minute scheduling changes were necessary.

Contents

Illustrations

Tables

Executive Summary

This project studied the sensitivity of two popular track inspection systems at points of interest (POIs)—points along the track that have deteriorated or may deteriorate in the near future. The first inspection system employs visual inspection[1] via a hi-rail vehicle[2], and the second system is an automated track geometry measurement system (TGMS) that outputs several metrics (collectively known as "track geometry"). These two systems were chosen due to their current and historical prevalence in the railroad industry. Two field data collection tests were conducted by QinetiQ North America and FRA's Office of Railroad Safety, with four participants per system for both field tests, while FRA's Office of Research, Development, and Technology conducted data analysis.

Signal detection theory quantified the systems' sensitivity to detect deteriorated track conditions and bias and decide whether there was a POI present. The deteriorated track conditions were grouped into a major bin (Maintenance Required, MR) as well as three sub-bins (Safety Critical, SC; Track Geometry Maintenance Required, TGMR; Track Geometry Safety Critical, TGSC), as shown in Figure 1 (page 11). Overall, there was no difference in sensitivity to POIs between the track inspectors and the TGMS system. However, track inspectors were more sensitive to POIs in the "Maintenance Required" bin while the TGMS system was more sensitive to conditions in the "Track Geometry Safety Critical" sub-bin. For the other sub-bins ("Safety Critical" and "Track Geometry Maintenance Required"), there was no statistically reliable differences in sensitivity between the two systems. In addition to the sensitivity of both systems, response bias was also studied. Overall, both the track inspectors and the TGMS were biased towards "no, there is no defect present."

The response bias observed in this report may be due to a number of factors. Response bias is influenced by the prior odds of observing a track deteriorated condition, which is low in this study and favored a bias to say "no." The values placed on detecting a deteriorated or anomalous track condition when there is none (a false alarm) versus missing such an anomalous condition that is present also determines response bias. If there is a policy (value) to avoid false alarms, a bias to say "no" will be established. A separate study to examine the influence of these values on track inspectors and TGMS operators may provide a better understanding of the results observed in this study.

---

[2] A hi-rail vehicle refers to a self-propelled vehicle that is manufactured to meet Federal Motor Vehicle Safety Standards and is equipped with retractable flanged wheels so that the vehicle can legally be used on both roads and rails. The name comes from combining "highway" with "rail." A common alternative spelling is "high-rail."

1. Introduction and Objectives

Since railroad track is complex and dynamic, maintaining railroad track to a specific standard is essential for the safe operation of trains throughout the rail network. The railroad industry currently monitors the condition of track with various manual and automated track inspection systems. Some of these inspections are mandated by federal regulations under Title 49, Section 213 of the Code of Federal Regulations (CFR). This section of the CFR is also known as the Track Safety Standards (TSS). Beyond the federally-mandated inspection requirements, frequently other inspections are voluntarily implemented by railroads to enhance the safety of their respective operations.

A long-standing goal of the railroad industry and the Federal Railroad Administration (FRA), which regulates railroad operations in the United States, is to develop optimal inspection strategies. To that end, probabilistic, risk-based models have been developed using mathematical techniques such as Markov chain methods and Monte Carlo methods. These optimal inspection strategies reduce the risk of a derailment due to poor track conditions, and enhance the safety of the railroad network.

This project examines two popular track inspection systems, which are used at points of interest (POIs) along track that is in deteriorated condition or may deteriorate in the near future. The first inspection system employs visual inspection and uses a hi-rail vehicle, while the second system is an automated track geometry measurement system (TGMS) that outputs several metrics known as “track geometry parameters” or simply “track geometry.” These two systems were chosen due to their current and historical prevalence in the railroad industry

However, the accuracy of a model’s output is heavily dependent on the accuracy of the inputs provided to the model. These inputs often include probabilities of detection of various inspection systems as well as degradation rates of various track anomalies. The industry continuously studies track degradation rates, largely through empirical data collection and analysis. However, the probability of detection of some popular track inspection systems has not been studied thoroughly to date.

Other manual and automated systems are currently used by the railroad industry, but they will not be covered in this report due to time and budget constraints, as well as the practical need for limiting the scope of the study. For example, in some cases, manual inspections are performed on foot instead of via hi-rail vehicles. On-foot inspections are common on corridors that experience high-traffic density, such as Amtrak’s Northeast Corridor, which runs between Washington, DC, and Boston, MA. However, the sensitivity of on-foot visual inspections will not be covered in this report. In addition, the sensitivity of automated non-track geometry systems will not be covered.

1.1 Organization of the Report

The report is organized into the following sections:

Section 2 – Presents background on federal regulations and current industry practice with regards to track inspection.

Section 3 – Discusses data collection activities.

Section 4 – Discusses the data analysis methodology and the results of the data analysis.

Section 5 – Provides the conclusions with general discussion points and hypotheses related to the results of the data analysis, and proposes future research work that may be pursued.

2. Background

2.1 FRA Regulations – Track Safety Standards

The TSS puts forth inspection requirements for track classes 1 through 5 in Subpart F, and portions of Subpart G includes requirements for track classes 6 through 9. The frequency and type of track inspection, as well as the record keeping requirements are described by the TSS. Visual inspections are required for all classes of track and can be performed on foot or from a hi-rail vehicle. Table 1 summarizes the visual inspection requirements for track classes 1 through 5.

Table 1. Visual inspection requirements outlined in the TSS

Class of Track	Type of Track	Required Frequency
Excepted track and Class 1, 2, and 3 track	Main track and sidings	Weekly with at least 3 calendar days' interval between inspections, or before use, if the track is used less than once a week, or twice weekly with at least 1 calendar day interval between inspections, if the track carries passenger trains or more than 10 million gross tons of traffic during the preceding calendar year.
Excepted track and Class 1, 2, and 3 track	Other than main track and sidings	Monthly with at least 20 calendar days' interval between inspections.
Class 4 and 5 track	………………	Twice weekly with at least 1 calendar day interval between inspections.

There are more details on the visual inspection requirements which are not covered in this report. For complete details, see Subpart F, Section 213.233 of the TSS (49 CFR 213.233) for track classes 1 through 5 and Subpart G, Section 213.365 of the TSS (49 CFR 213.365) for track classes 6 through 9.

In addition to the visual inspection requirements, the TSS mandates automated track inspection with a TGMS for higher classes of track, namely track classes 6 through 9. Requirements for automated track geometry inspection can be found in Section 213.333 of Subpart G in the TSS. Other automated systems besides track geometry are prescribed in the TSS, such as internal rail inspection. However, these systems were not included as part of this study and will not be discussed further.

2.2 Current Industry Track Inspection Practices

Current industry practice with regard to track inspection is partially governed by federal regulations put forth in the TSS, but since the TSS prescribes *minimum safety standards*, some railroads enhance safety and efficiency by adopting more stringent standards. As a result, these railroads often employ additional inspection procedures beyond those mandated by the TSS. Using automated track geometry cars on lower classes of track, namely track classes 5 and lower, may be the most obvious example of adopting of additional inspection systems and frequencies besides those prescribed in the TSS. All Class 1 railroads[3] make use of geometry cars even though the great majority of the track owned by Class 1 railroads is track class 5 or lower.

2.3 Track Geometry and TGMS

The term "track geometry" covers the vertical and lateral deviations of each of the rails as well as the gage and cross-level measurements, which are, respectively, the horizontal and vertical relationships between the two rail heads. Maintaining proper track geometry is essential in order to maintain acceptable ride quality and prevent derailments. A TGMS is a system that is capable of measuring such track geometry parameters as gage, cross-level, alinement[4], and surface (also known as "profile") deviations at a given time or spatial interval. A TGMS is mounted on either a full-size railcar or a hi-rail vehicle, but using a TGMS on a full-size railcar is preferred since the track geometry is measured under load when a full-size railcar is utilized.

TGMS are widely utilized in the railroad industry, largely on a voluntary basis. There are several suppliers of geometry systems throughout the industry, and they each use unique technologies. Some systems are based on integrating and filtering of inertial sensor data, while others rely on mechanical means to measure mid-chord offsets (MCOs) then convert the MCOs to track geometry parameters such as alinement and profile with various algorithms.

Additional technical details about geometry systems can be found in the literature. Most importantly, TGMS are complex, automated systems and there is not a single correct or accepted way to implement them. Each supplier has their own "mix" of software algorithms and hardware technologies in place. Oftentimes, the internal workings of the hardware and software are proprietary and the exact science and engineering behind a respective system may not be publicly available.

This study only made use of a TGMS from a single supplier. Thus, it does not compare the variance between different suppliers of TGMS systems.

---

[3] The term "Class 1 railroad" refers to a large railroad company. Class 1 status depends on operating revenue. In the United States, there are seven official Class 1 railroads: Burlington Northern Santa Fe (BNSF) Railway Company; Canadian National (CN) Railway Company; Canadian Pacific (CP) Railway; CSX Transportation, Inc.; Kansas City Southern (KCS) Railway Company; Norfolk Southern (NS) Railway Company; and Union Pacific (UP) Railroad Company. The National Railroad Passenger Corporation (Amtrak), although not technically a Class 1 railroad, is oftentimes unofficially referred to as such. Note that the term "Class 1 railroad" does *not* have any relation to track classes 1 through 9 in the TSS.

[4] The TSS uses this spelling, but the more common spelling is "alignment".

3. Data Collection Preparation and Execution

The study included one pilot field test and two official data collection field tests. While the pilot field test was conducted on a segment of the Bay Line Railroad near Panama City Beach, FL.[5], the first and second official field tests occurred on segments of track in Virginia (known as RR1 in this study) and Texas (known as RR2). This section provides details on the equipment used, the test participants, and the ground verification process. Exact timelines and test plans can be found in Appendices A and B.

3.1 Test Equipment

The following outlines the primary equipment required for the study.

Visual Inspection Testing:

- A hi-railer which served as the platform for participating FRA track inspectors to perform their respective inspections. The hi-railer also served as a tool for traversing the test zone in an expedient manner during the ground verification process.
- A customized push button data acquisition system.

TGMS Inspection Testing:

- A TGMS system mounted on a second hi-rail vehicle.

Ground Truth Verification:

- Tools such as a string lining kit, a gage level board, an 18-inch straight edge, and a GPS unit for accurate positioning.

3.2 Test Participants

The following groups participated in this study:

1) Track inspectors (TI system or observers) who performed a visual inspection of the track via a hi-rail vehicle.
2) Human exception editors who processed TGMS data. These exception editors reviewed the exception reports from the TGMS to identify and filter out potential false alarms. This combination of TGMS data and exception editors will be referred to as the edited track geometry (ETG) system or observers.

3.2.1 Track Inspectors

Test participants to perform visual inspections were recruited from among the pool of FRA TIs. Random sampling was not logistically feasible so convenience sampling was used to recruit FRA TI participants for the study.

---

[5] This field test served as a pilot field, and as such, data from this field test is not included in the data analysis section of this report.

The study was conducted as part of their normal working hours. Therefore, no special compensation for participation was provided, and travel for the participants was borne directly by FRA. There were four TI observers per field test.

Upon arrival at the field test location, each TI observer was invited to read the consent form (see Appendix C). It was emphasized that participation is voluntary and can be ended at any time at the participant's discretion. Field test organizers explained that all participant specific data collected will be utilized for study purposes only and will only be shared either anonymously or in an aggregated manner. It was further emphasized that each inspector's respective performance in the study would not affect his or her job in any way. TI participants were asked if they have any specific safety or privacy concerns. Once all questions were resolved and the participant was satisfied, he / she signed the consent form.

After the consent forms were signed, the participants were given test instructions. TIs participating in the study were asked to locate POIs (anomalous or deteriorated track conditions). A POI may be a track condition that is not in compliance with the TSS or any type of track condition that, when performing a typical FRA compliance inspection, would cause the hi-rail vehicle to be stopped to perform a ground inspection. The segments of track used for the first and second field tests (RR1 and RR2, respectively) are posted as Class 3 track. However, in both cases, participants inspected the track to Class 4 standards, which increased the sample size of the POIs.

The TI observers recorded all POIs by pressing a handheld button attached to a custom data acquisition system then concisely describe why they pressed the button. Each participant was given instructions and a brief demonstration of the data acquisition system and the attached button. After this demonstration, each participant conducted an inspection of the entire test zone, one inspector at a time.

The TI observers conducted their inspection from the front passenger seat of the hi-rail vehicle. Despite not having access to the vehicle's brake and acceleration pedals, the TIs were allowed to direct the driver to speed up or slow down. In addition, the TIs could stop the vehicle and back it up. It was emphasized that they may travel as slowly as they need to do a thorough and fastidious inspection, but they were encouraged to maintain a minimum speed of five miles per hour. The maximum hi-railer speed was limited by the speed of the class of track or by the railroad's operating procedures. For both field tests, typical hi-railer speeds would not exceed 25 miles per hour.

The TI system of inspection is typically a two stage process in which the TI observer initially inspects the track from a hi-rail vehicle and then leaves the hi-rail vehicle to conduct a more thorough ground inspection when they spot a potential anomalous track condition from within the hi-railer. In the second stage, the TI observer may reconsider their initial suspicion that the location was suffering from deterioration, and therefore, they could choose not to include the location in a final list of deteriorated track locations. Due to time constraints and overall logistical considerations, TI observers participating in the study were encouraged not to depart the hi-rail vehicle. Therefore, this study only tested the first phase of the TI system, where the inspector makes an initial detection of a potential deteriorated track condition from the hi-rail vehicle.

3.2.2 TGMS System and Exception Editors

Four data collection runs were conducted with the TGMS system at both the RR1 and RR2 test sites. The TGMS travelled at the maximum allowable hi-railer speed, which was 25 miles per hour for both test sites.

After the on-site data collection was completed, the TGMS supplier performed human-based editing of the system's exception locations. The editing process eliminates the false alarms output by the TGMS due to sensor malfunctions or general system anomalies. The TGMS exception editors view the data trace output near exception locations then de to keep or discard that exception location based on the data plot as well as their knowledge of the TGMS system and its strengths and weaknesses. The resulting combination of TGMS data with exception editors shall be referred to as edited track geometry (ETG).

The TGMS exception editors were provided by the supplier of the TGMS system, and the study was conducted during their normal working hours. Therefore, no special compensation for participation was provided. The filtering process was conducted at the supplier's office location after the field test data was collected, and as such, travel costs were not incurred. All exception editors were experts with several years of experience.

As with the TI participants, each ETG participant was asked to read the consent form. It was emphasized that participation in the study was voluntary and can be ended at any time at the participant's discretion. It was explained that all participant specific data collected would only be utilized for study purposes and would not be shared with their employer, except in an aggregate manner or anonymously. It was also explained that their performance in the study would not affect their employment in any way. Each participant was asked if they have any specific safety or privacy concerns. Once all questions were resolved and the participant was satisfied, the consent form was signed.

Each editor then filtered the TGMS exception reports as they would normally. They were presented with a playback of the TGMS data feed in the form of strip charts and asked to filter the exceptions using their normal decision making processes to Class 4 track standards. This process was completed once for each editor. Each editor viewed data from only a single geometry run. There were four editors in total, so each editor viewed a single run from the RR1 field test as well as a single run from the RR2 field test.

For all of the TGMS data collection runs during this research effort, the track geometry system operators did not confirm/reject potential track geometry defects during the original collection of the data. All data was reviewed and "edited" by trained, experienced track geometry system operators (exception editors) after the survey. These exception editors were prohibited from reviewing data collected during a survey they participated in; i.e. exception editors had no prior experience with the specific track that data they were tasked with reviewing was collected on.

As part of the normal collection of track geometry data, the survey vehicle's Forward Observer (FO) will mark the presence of crossings, switches and other track features within the data in accordance with standard operations; this process was followed during the course of this research effort. In addition, the track geometry measurement system employed during this research employed an Automatic Location Detector (ALD) that returns a "non-zero" signal when it passes over switch rails within the track gauge thus recording the location of switches; this signal is available on both manned and unmanned geometry measurement systems employed by FRA at the time this research was conducted. During the data review process, exception editors were provided only with track data which was collected as part of the survey, including assets marked by the FO and the presence of a switch as indicated by the ALD signal. No outside information—such as track charts, satellite imagery or speed tables—was made available to the exception editors during their review/editing process.

3.3 Ground Verification and Data Binning

The ground verification phase was a crucial aspect of this study. For both the RR1 and RR2 field tests, an independent consultant with an extensive background in track inspection was used to conduct the ground verification. The same consultant was used at both field test locations in order to achieve consistency and avoid introducing another variable into the study.

For both field tests, the primary, comprehensive ground verification was performed following all the track inspection runs by the TI observers and the TGMS system. However, it was determined that to remove potential bias and increase sensitivity, having a brief pre-validation of the track in addition to the primary, comprehensive ground verification would be the best way to perform the overall ground verification process. Therefore, the consultant performed his preliminary validations on both the RR1 and RR2 field tests on the first day of each respective field test, prior to any track inspector or TGMS data collection. This preliminary, unbiased field validation allowed the independent consultant to have an unclouded, unbiased view of the various conditions that existed in the test zone. The independent consultant documented any obvious POIs and systemic problems.

The more comprehensive ground verification phase for the RR1 and RR2 field tests each took two days to complete. This comprehensive ground verification process was conducted after all the track inspector and TGMS data collection runs. All those involved in the study were welcomed to participate in aiding the consultant in his validation of POIs found throughout the previous days of data collection.

Using a custom designed GPS system that integrated real world location of the POIs into a Google Maps display, the independent consultant was able to travel to the precise locations identified by the TI observers and the TGMS. At each location, the independent consultant would exit the vehicle to inspect the locations on foot (unless it was obvious from within the hi-rail vehicle that a POI was present). Any POIs that the independent consultant found either in his first or second validation that were not found by either the track inspectors or the TGMS were considered to be “Misses” (also known as “False Negatives”) within the purview of signal detection theory analysis.

There were two criteria[6] or threshold levels that the independent consultant used: a safety level threshold and a maintenance level threshold. For each POI indicated by the TI observers and the TGMS, the consultant would rate whether it was a hit (also known as a “true positive”) for the safety threshold as well as whether it was a hit for the maintenance threshold. If it was a hit for the safety threshold, then it was inherently considered to be a hit for the maintenance threshold. In other words, the locations on the track that were a safety concern were a subset of the locations that were a maintenance concern. The main bin of all maintenance locations will be

---

[6] A threshold level is referred to as a “criterion” in signal detection theory terminology.

referred to as "maintenance required" (MR) since maintenance is required at those locations either immediately or in the foreseeable future. The sub-bin of safety critical exceptions shall be referred to as the "safety critical" (SC) sub-bin.

For analysis purposes, the MR bin was further divided into a total of three sub-bins (including the SC sub-bin). The track geometry maintenance required (TGMR) sub-bin contained maintenance conditions in which track geometry played at least a partial role in the deteriorated condition. For example, an area of the track that required maintenance due to some track geometry deviations in combination with fouled ballast or degraded ties would be placed in this sub-bin. Finally, the track geometry safety critical (TGSC) sub-bin contained safety critical conditions that were at least partially due to track geometry. The TGSC sub-bin represents the intersection of the SC and TGMR sub-bins as shown in Figure 1.

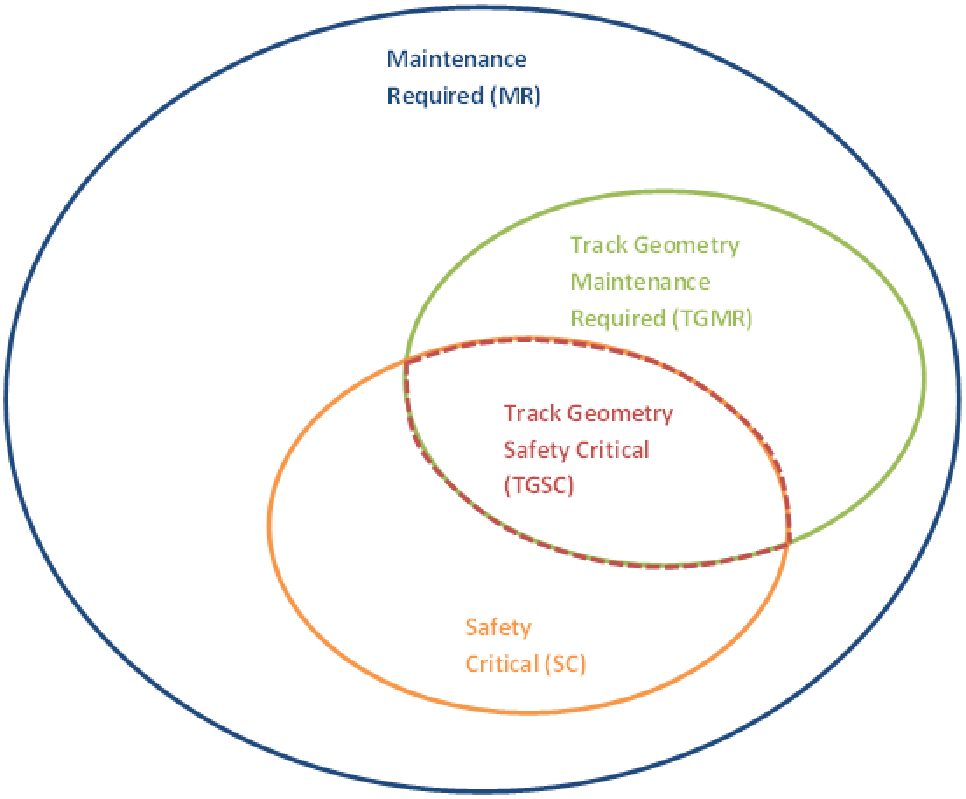


Figure 1. Venn diagram showing the primary MR bin in addition to the three sub-bins (SC, TGMR, and TGSC)

4. Data Analysis

4.1 Signal Detection Theory Primer

Signal detection theory (SDT) was used to quantify the sensitivity of human track inspection via a hi-rail vehicle as well as the sensitivity of the edited TGMS data. In SDT, a system (TIs or ETG) make decisions about the presence of a signal (a location in the track that is a safety or maintenance concern) in a background of noise, such as electrical noise in an electronic system or visual noise in the form of track conditions that may be present but do not rise to the level of a safety concern or a maintenance concern. In SDT, it is often assumed that both signal and noise can be represented by overlapping normal distributions[7] (Green Swets, 1966; Macmillan & Creelman, 2005), as shown in Figure 2.

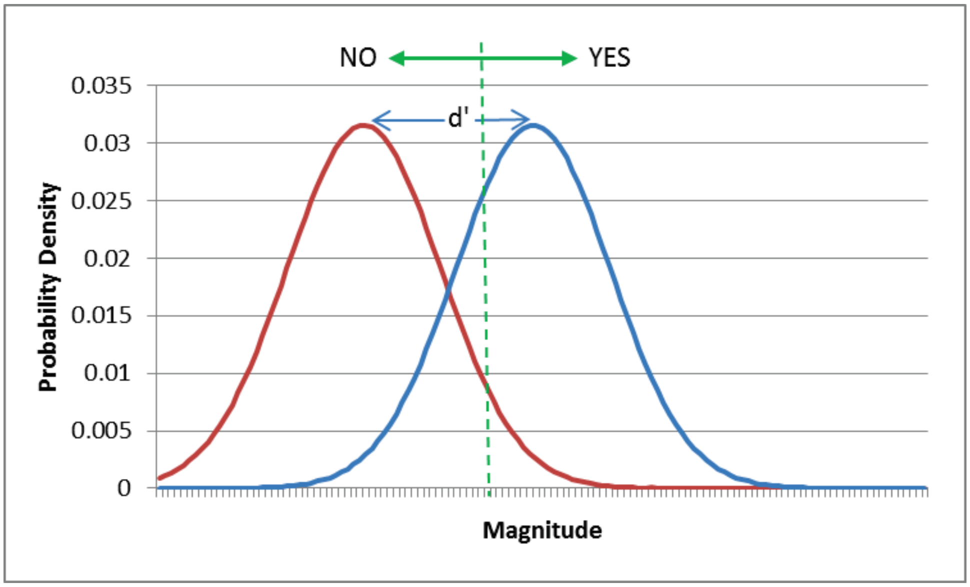


Figure 2. Example overlap of a normal noise (red) distribution and a normal signal (blue) distribution

The system under investigation sets a criterion line along the decision axis (e.g., the magnitude of the anomalous track condition) which divides the overlapping distributions into four decision outcomes. If an event magnitude falls to the right of the criterion, the system responds "yes, there is an anomalous track condition", and if an event magnitude falls to the left of the criterion, the system responds "no, there is no anomalous track condition". If the event was a signal, a "yes" response represents a "Hit" (also known as a "True Positive"). If the event was noise, a "yes" response represents a "False Alarm" (also known as a "False Positive"). A "Miss" (also known as a "False Negative") is a "no" response to a signal event, and a "Correct Rejection" (also known as a "True Negative") is a "no" response to a noise event. The possible decision outcomes are summarized in Table 2.

---

[7] This assumption is not necessary to SDT. Other probability distributions can be used, or data can be analyzed in the absence of knowledge concerning the underlying distributions (Macmillan Creelman, 2005).

Table 2. Possible SDT outcomes

		State of the track	
		POI	No POI
Inspection Response	Yes	Hit	False Alarm
	No	Miss	Correct Rejection

It should be noted that Hits and Misses both come from the signal distribution, and False Alarms (FAs) and Correct Rejections (CRs) both come from the noise distribution. Consequently, the probability of a Hit (*p*(Hit)) and the probability of a Miss (*p*(Miss)) sum to 1:

$$p(\text{Hit}) + p(\text{Miss}) = 1 \tag{1}$$

With regard to Figure 2, *p*(Hit) was represented by the area under the signal distribution to the right of the criterion. Similarly, the probability of a False Alarm (*p*(FA)) and the probability of a Correct Rejection (*p*(CR)) sum to 1:

$$p(\text{FA}) + p(\text{CR}) = 1 \tag{2}$$

With regard to Figure 2, *p*(FA) was represented by the area under the noise distribution to the right of the criterion. As a result of the two unity equations above, all the information about the underlying response probabilities can be captured by knowing *p*(Hit) and *p*(FA).

SDT was used as an analytic tool because it provides a way to examine a system's decisions with respect to the detectability of anomalous or deteriorated track conditions (known as "sensitivity") and to attitudinal or motivational factors that may influence a system's criteria for judgment (known as "bias"). SDT describes the system's ability to detect a signal in a background of noise as a discrete choice task. The separation of the noise and signal distributions determines sensitivity. If the underlying distributions are normal and have equal variance, as shown in Figure 2, then the difference between the means of the noise and signal distributions is called *d'*[8] and can be calculated from the difference of the normal transforms of *p*(Hit) and *p*(FA):

$$d' = z(\text{Hit}) - z(\text{FA}) \tag{3}$$

Response bias (commonly referred to as β if the distributions are normal) was determined by the placement of the criterion along the decision axis. Bias reflects a system's tendency to say "yes" or "no", and it is depicted by the dashed green vertical line in Figure 2. A shift to the right demonstrates conservative behavior; that is, the system was more likely to respond "no", which decreases the number of Hits but also the number of FAs. Response bias sets the location of the criterion and causes *p*(Hit) and *p*(FA) to co-vary while *d'* remains constant.

Bias is influenced by the value of decision outcomes. For example, a Miss may result in an accident that results in damage to track and equipment and potential injuries, while a FA could cause a delay in rail traffic and result in a certain amount of lost revenue. Bias was also influenced by the prior odds, which were defined as the probability of noise (*p*(Noise)) divided by the probability of signal (*p*(Signal)):

---

[8] Pronounced "dee prime".

$$\text{Prior Odds} = \frac{p(\text{Noise})}{p(\text{Signal})} \tag{4}$$

If signals are rare events, the criterion shifts to the right, and the system would be more likely to respond “no”, which decreases the number of Hits and the number of FAs. Bias is related to the prior odds and the value of decision outcomes in Table 2 as follows:

$$\beta = \left(\frac{V(\text{CR}) + V(\text{FA})}{V(\text{Hit}) + V(\text{Miss})} \right) \left(\frac{p(\text{Noise})}{p(\text{Signal})} \right) \tag{5}$$

Table 3 shows the payoff matrix corresponding to the above equation.

Table 3. Payoff matrix for SDT outcomes

		State of the track	
		POI	No POI
Inspection Response	Yes	Value of Hit, V(Hit)	Value of False Alarms, V(FA)
	No	Value of Misses, V(Miss)	Value of Correct Rejections, V(CR)

Figure 3 shows examples of receiver operating characteristic (ROC) curves. ROC curves show the probability of a Hit (*p*(Hit)) as a function of the probability of a FA (*p*(FA)). When *d’* equals zero, the ability to detect an event is null (also known as the major diagonal). The isosensitivity curve (a curve of constant sensitivity showing *d’* equal to 2.2 corresponds to the *p*(Hit) and *p*(FA) values that would be generated by the normal distributions depicted in Figure 2 as the criterion is moved from left to right. Figure 3 also shows the isobias line (a line of constant bias) for bias equal to zero (also known as the minor diagonal). This corresponds to having the criterion set at the intersection of the noise and signal distributions. Points on isosensitivity curves above the minor diagonal correspond to a bias to say “yes”, and points below the minor diagonal correspond to the bias to say “no”.

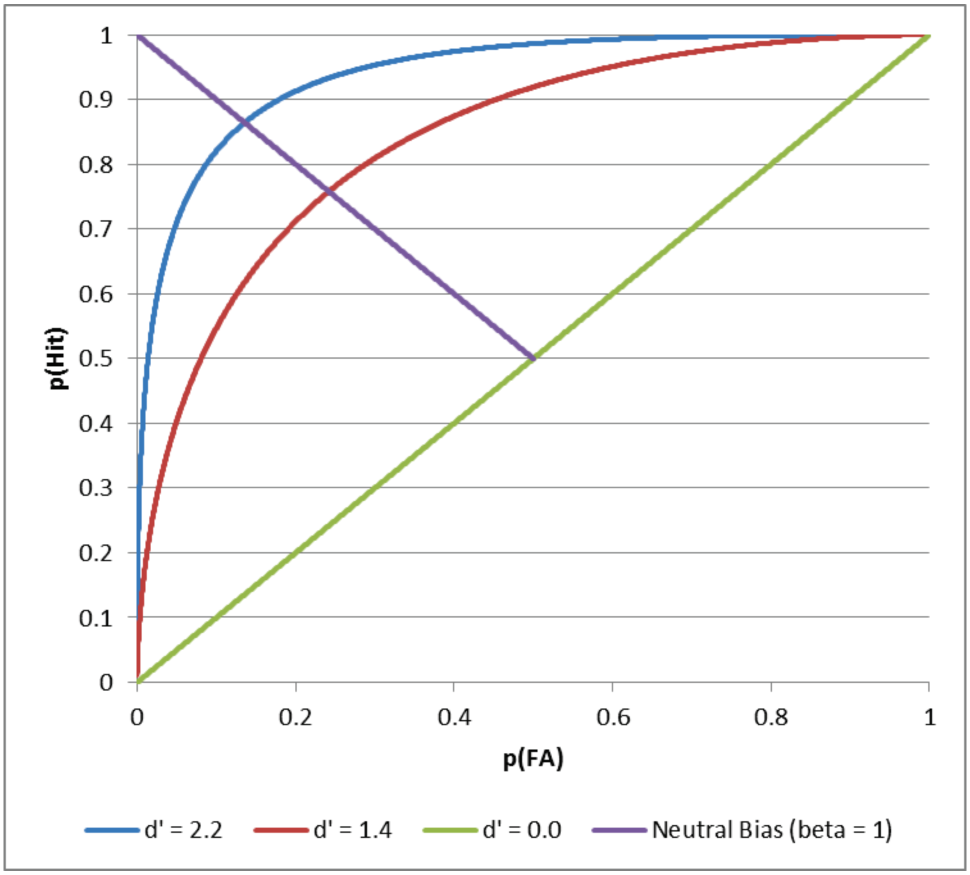


Figure 3. Three example ROC curves

If the underlying distributions in Figure 3 are both normal and have equal variances, it is expected that bias and sensitivity are independent (not correlated). A major advantage of the use of SDT in detection tasks is this ability to assess sensitivity independently of response bias. A diagnostic test for normal distributions with equal variances is to plot the ROC in normal-normal coordinates (see Egan, 1975). The resulting isosensitivity curves are straight lines with unit slopes. A difference in slope between conditions indicates that the distributions have different variances, while non-linear isosensitivity curves indicate the distributions are not normal. Figure 4 shows the ROCs from Figure 3 in normal-normal coordinates.

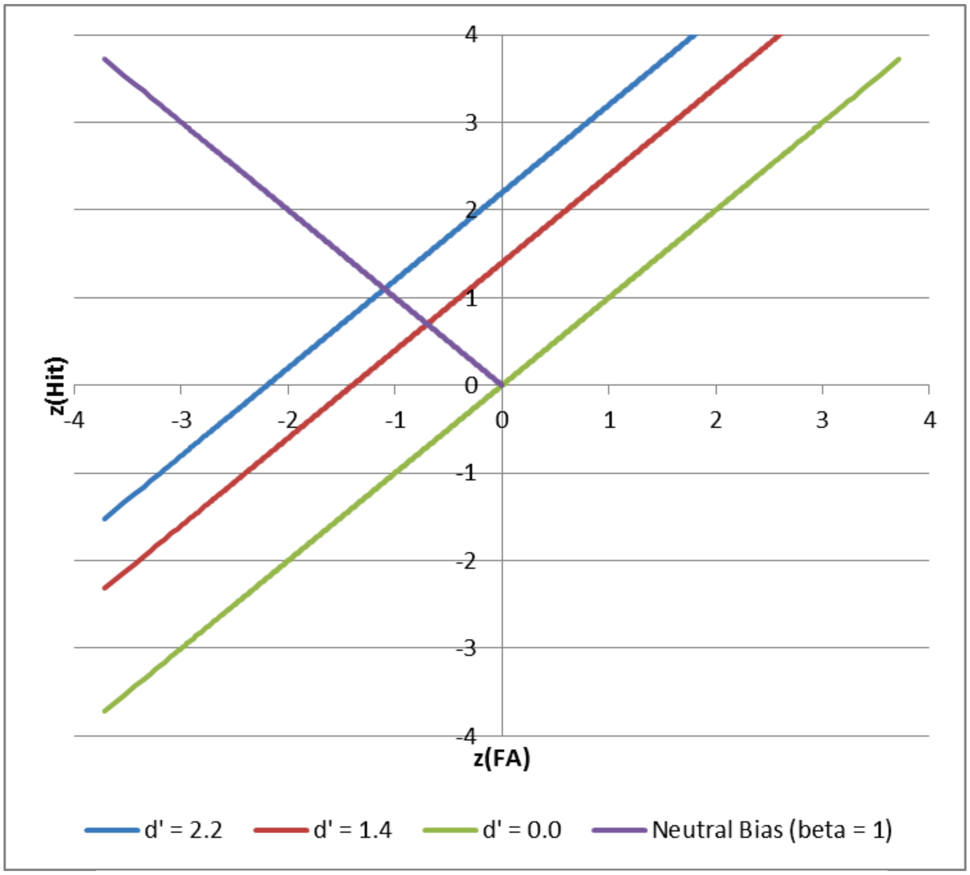


Figure 4. ROC curves from Figure 3 in normal-normal coordinates

If the distributions are not equal variance normal and the distribution type is unknown, nonparametric indices of sensitivity and bias can be used to analyze the data (Macmillan and Creelman, 2005). The nonparametric sensitivity parameter is referred to as *A'* and its value can be determined by the following formula:

$$A \quad \frac{(p(\mathrm{Hit}) - p(\mathrm{FA}))(1 + p(\mathrm{Hit}) - p(\mathrm{FA}))}{4p(\mathrm{Hit})(1 - p(\mathrm{FA}))} + 0.5 \tag{6}$$

The nonparametric bias parameter is referred to as *B''* and its value can be determined by the following formula:

$$B \quad \frac{p(\mathrm{Hit})(1 - p(\mathrm{Hit})) - p(\mathrm{FA})(1 - p(\mathrm{FA}))}{p(\mathrm{Hit})(1 - p(\mathrm{Hit})) + p(\mathrm{FA})(1 - p(\mathrm{FA}))} \tag{7}$$

As will be seen in the following section, the data collected did not meet normality assumptions, and the data analysis consequently made use of the *A'* and *B''* parameters for sensitivity and bias, respectively.

4.2 Results

Data was collected on two different field tests (RR1 and RR2, respectively). Two systems were being tested. The first system was human visual inspection via a hi-rail vehicle (TI system), and the second system was an automated track geometry system with human editing of exception data (ETG system). There were four observers per system.

Both RR1 and RR2 were segments of single track. This is important to point out since the TSS allows inspection of multiple tracks by a single TI observer. This study limited the TI observer to inspecting a single track at a time.

With SDT, for a given length of track, or area of possible POI, one and only one of four conditions can be present. Either there has been a CR, a Miss, or a Hit, or a FA for this section. Multiple Hits and Misses cannot occur on the same area of track being analyzed. For this reason, the test zone lengths on both the RR1 and RR2 field tests were broken up into 160 foot sections.

Typically, each section of jointed rail is 39 feet, so 160 feet represents approximately four jointed rail lengths. It was felt that this was a good section length since the participants in the study might not press the data acquisition button at exactly the time that the hi-railer passed over their POI. Hi-railer speeds approached 25 miles per hour at times, which is equivalent to about 37 feet per second. So a delay in pushing the button by two seconds, for example, could result in a 74 foot offset. In addition, GPS positioning error is always present to a certain degree. Since setting the segment length too small was not desirable, 160-foot segment lengths were chosen.

4.2.1 Normality Tests

Values of *p*(Hit) and *p*(FA) were calculated from the data provided in Appendix D and used to determine values of *d'* and β to test if the typical normality assumptions of most SDT analyses were applicable. Several factors indicated that SDT with assumptions of normal distributions with equal variance was not appropriate for analyzing this data set. First, there is a reliable, statistically significant correlation between *d'* and β, which is shown in Figure 5.

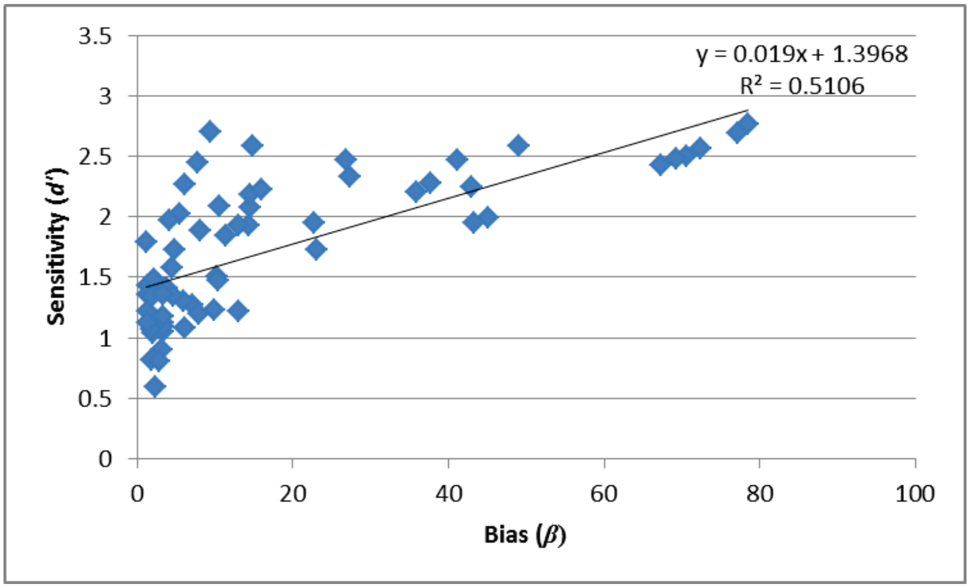


Figure 5. Scatterplot of *d'* and β

Second, the ROC for all the data indicates that the variance of the two observer groups (TI observers and ETG observers) is different. Figure 6 shows the ROC of the data plotted on normal-normal axes. The slope for the ETG system is 0.4 which is only about half that of the TI system which has a slope of 0.73. This indicates a disparity in variance between the two systems.

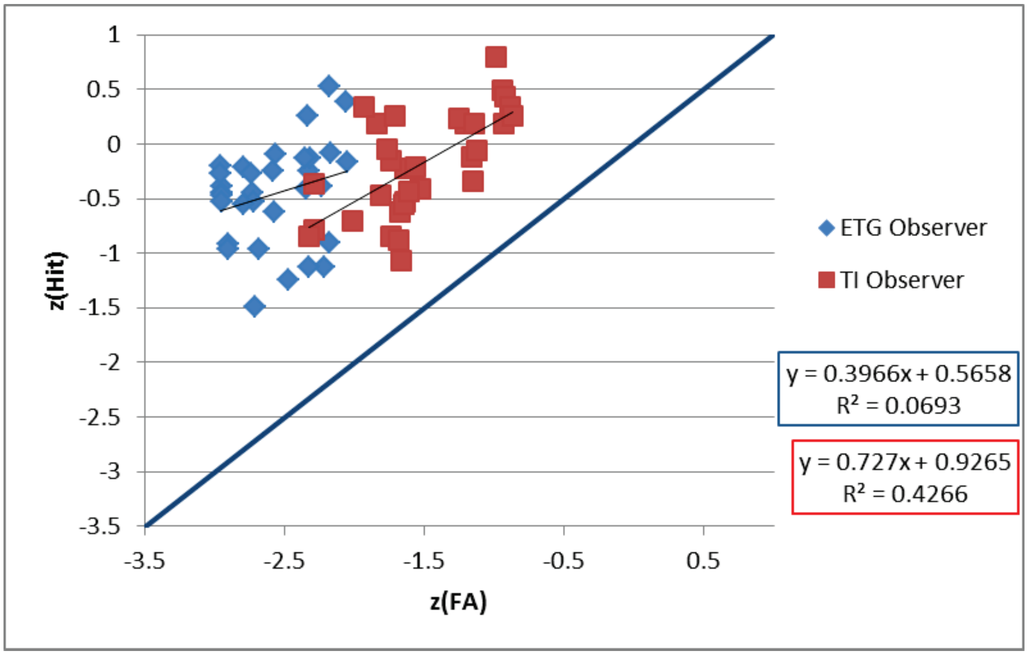


Figure 6. *z*(Hit) vs. *z*(FA) for all participants (TIs and ETGs) and observation conditions (MR, SC, TGMR, and TGSC; see Figure 1)

Third, the sampling procedure for the data resembles a Poisson process in which a segment of track is inspected for detects (see Daniel, 1978) and it could result in underlying Poisson distributions for signal and noise for the ETG and TI observer groups. This could be the source of the differences in variance suggested by Figure 6 since the mean and variance of a Poisson distribution are equal, unlike the normal distribution in which the mean and variance are independent (Daniel, 1978). Since there is insufficient data to estimate parameters for Poisson distributions, the nonparametric indices for sensitivity (*A'*) and bias (*B''*) were used in subsequent analyses.

4.2.2 ROC Plots

A plot of *p*(Hit) as a function of *p*(FA) is shown for all the data in Figure 7. Both ETG and TI observers have a bias to say “no” as indicated by the distribution of the points below the minor diagonal, which represents zero (or neutral) bias. TI observers appear to have a smaller bias to say “no”. The sensitivity (*A'*) of the TI system and ETG system will be discussed in more detail in the next section.

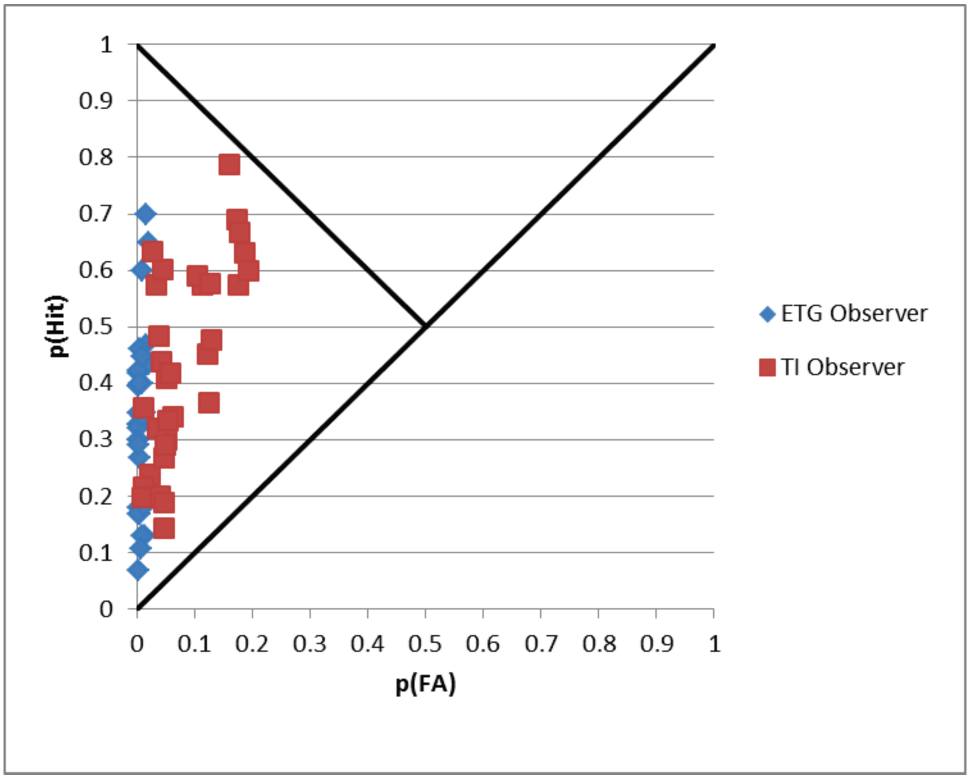


Figure 7. ROC plot of *p*(Hit) vs. *p*(FA) for TIs and ETGs

The values of *p*(Hit) have the greatest variability (variance = 0.029) while *p*(FA) varies much less (variance = 0.003). This difference is statistically reliable. The variance in *p*(FA) is much less for the ETG system (variance = 3.16 x 10-5) than for the TI system (variance = 0.003), and again that difference is statistically reliable. The pattern for the ETG observers suggests that this group is setting the criterion to maximize the Hit rate for a fixed FA rate. Observers who do this are known as Neyman-Pearson observers (Egan, 1975).

Figure 8 shows the mean values of *p*(Hit) and *p*(FA) for the ETG and TI systems. The error bars indicate 95% confidence intervals. The means for *p*(Hit) agree with the visual inspection of Figure 7 that there is no significant difference for *p*(Hit) between ETG and TI. There are statistically reliable differences between *p*(Hit) and *p*(FA) and between *p*(FA) for ETG and TI. The ETG system has a lower *p*(FA) than the TI system.

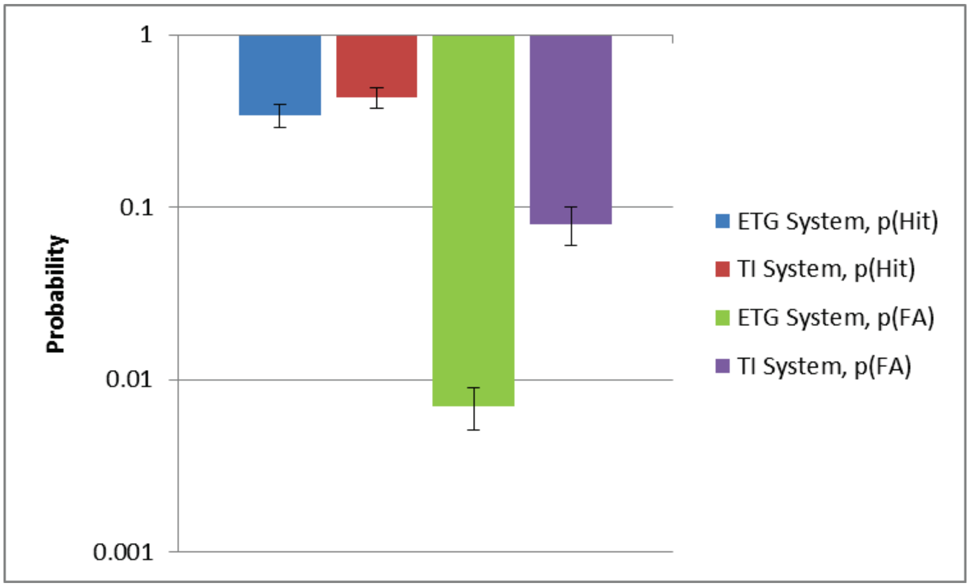


Figure 8. Mean *p*(Hit) and *p*(FA) for TIs and ETGs

Table 4 shows the mean values of *p*(Hit), *p*(Miss), *p*(FA), and *p*(CR) for both observer systems and all four observation conditions. Of particular note is the high *p*(Miss) rate for both the ETG and TI system across all observation conditions. The values of *p*(Miss) range from a low of 0.499 to a high of 0.858.

Table 4. Mean values of *p*(Hit), *p*(Miss), *p*(FA), and *p*(CR)

		Maintenance Required (MR)	**Safety Critical (SC)**	**Track Geometry Maintenance Required (TGMR)**	**Track Geometry Safety Critical (TGSC)**
Edited Track Geometry (ETG)	*p*(Hit)	0.142	0.364	0.359	0.501
	p(Miss)	0.858	0.636	0.641	0.499
	p(FA)	0.007	0.009	0.005	0.008
	p(CR)	0.993	0.991	0.995	0.992
Track Inspector (TI)	*p*(Hit)	0.395	0.445	0.463	0.439
	p(Miss)	0.605	0.555	0.537	0.561
	p(FA)	0.027	0.097	0.091	0.105
	p(CR)	0.973	0.903	0.909	0.895

4.2.3 Detection of Deteriorated Track Conditions (Sensitivity)

The nonparametric measure of sensitivity *A'* is the measure of sensitivity used in this analysis. *A'* can have values between 0.5 (indicating no ability to detect the signal) and 1.0 (indicating perfect ability to detect the signal).

Figure 9 shows the mean *A'* values for both field test locations. Although RR2 has a higher *A'* than RR1, the overlapping 95% conference intervals indicate that this difference is not statistically reliable. Consequently, data analyses of *A'* are collapsed across both field test locations.

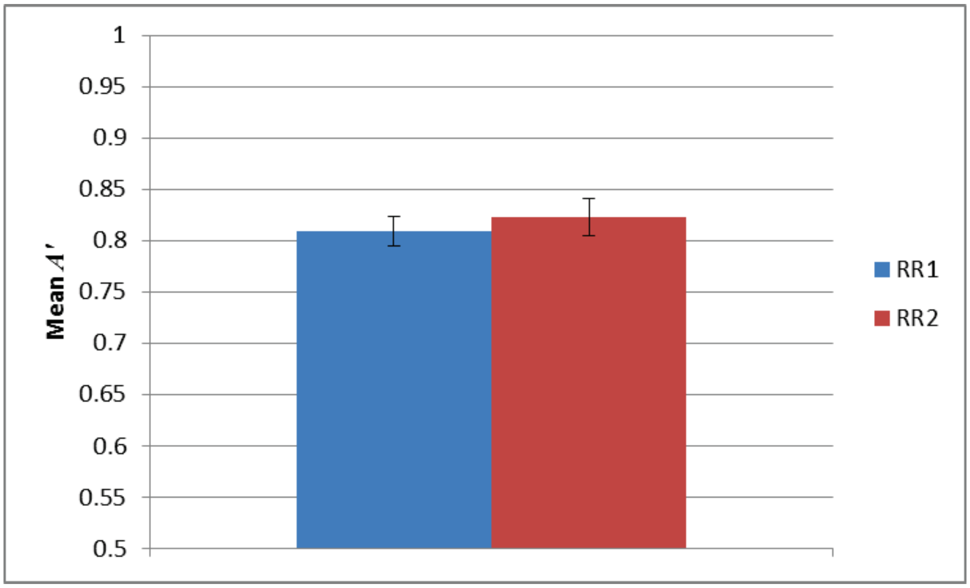


Figure 9. Mean *A'* and 95% confidence intervals for RR1 and RR2 across all four observation conditions and both observer groups

The sensitivity of the ETG and TI systems is shown in Figure 10. Mean *A'* is higher for the ETG system. Once again, the overlapping 95% confidence intervals indicate that this difference is not statistically reliable, which agrees with the visual inspection of the ROC plot in Figure 7. This study did not include a power calculation (see Cohen, 1988) prior to data collection to determine if the sample size was sufficient to detect a difference between the two groups of observers. This issue will be examined in the discussion and conclusions section (Section 5).

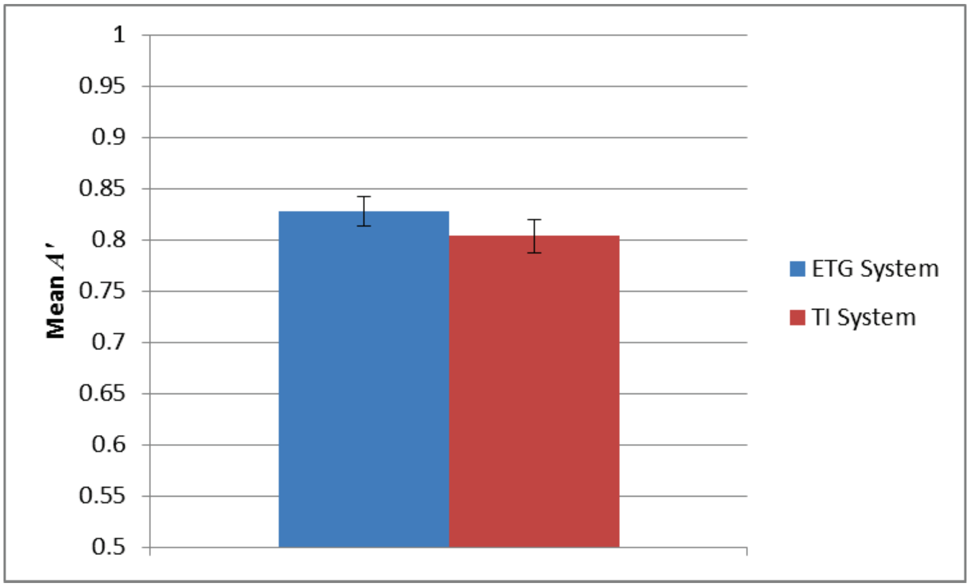


Figure 10. Mean *A'* and 95% confidence intervals for ETGs and TIs across all four observation conditions

Figure 10 shows the entire range of possible *A'* values (0.5 to 1.0) along the vertical axis, but it is hard to see the overlapping 95% confidence intervals in this figure. Therefore, a zoomed-in view of Figure 10 is shown in Figure 11 to make it easier to see the overlapping confidence intervals.

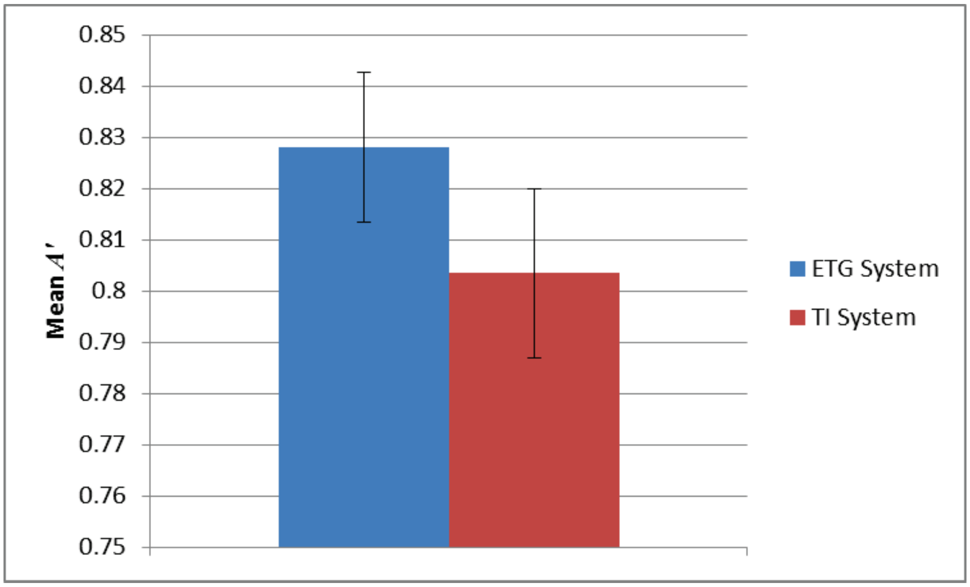


Figure 11. Zoomed-in view of Figure 10

Figure 12 shows the mean *A'* values for the four observation conditions (All Maintenance, All Safety, Track Geometry Maintenance, and the Track Geometry Safety). Given the fact that *A'* values can range from 0.5 to 1.0, it can be seen that there is not much difference in *A'* values for the four observation conditions.

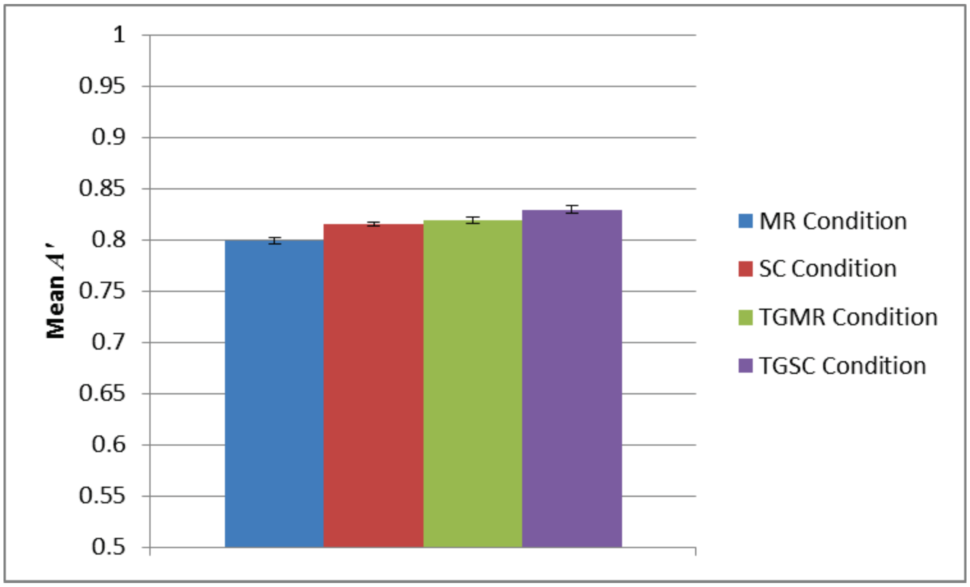


Figure 12. Mean *A'* and 95% confidence intervals for all four observation conditions from both observer groups on both railroads

Figure 13 shows a zoomed-in view of Figure 12 which makes it easier to see if the 95% confidence intervals are overlapping. The MR condition has the lowest detection and the TGSC condition has the highest detection rate overall. The SC and TGMR conditions fall between the other conditions and are not reliably different from each other. The MR and TGSC conditions are statistically different from each other and from the other two conditions.

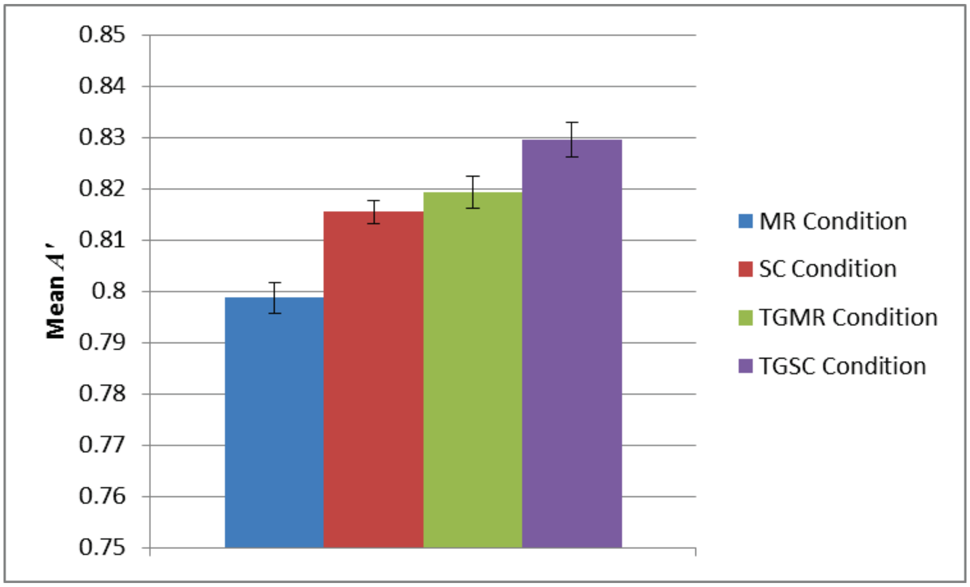


Figure 13. Zoomed-in view of Figure 12

Figure 14 explores the possibility that there is an interaction in sensitivity between the observation conditions and the observer system. Here we see a complicated interaction between these two variables. For the Track Inspectors *A'* does not appear to vary significantly across observation conditions as indicated by the overlapping 95% confidence intervals. However, there is a clear, statistically reliable increase in detection of POIs for the ETG system as one proceeds from the MR condition to the TGSC condition.

The ETG and TI systems show clear differences in detection with regard to the MR and TGSC observation conditions. TIs do reliably better at detection of a MR condition, but the ETG system is reliably better at detecting a TGSC condition. The groups are not reliably different in the other conditions.

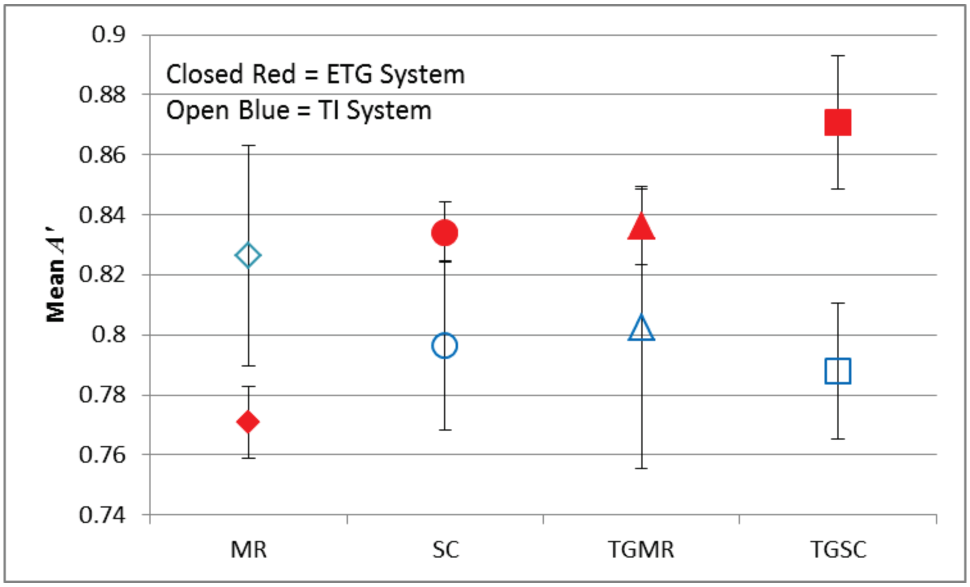


Figure 14. Mean *A'* and 95% confidence intervals for ETGs and TIs for all four observation conditions

4.2.4 Response Bias

B'', the measure of response bias used here, can have values that range from -1 (extreme bias to say "yes") to +1 (extreme bias to say "no"). If *B''* equals zero, there is no bias.

It was previously determined that *d'* (sensitivity) was reliably correlated with β (response bias), which is a violation of the assumptions of SDT (see Figure 5) that sensitivity and bias are independent. Figure 15 demonstrates that *A'* is independent of *B''*. The correlation between *A'* and *B''* is 0.23 ($p > 0.05$) which is not statistically reliable. Consequently, *a'* and *B''* provide independent information about the track inspection task.

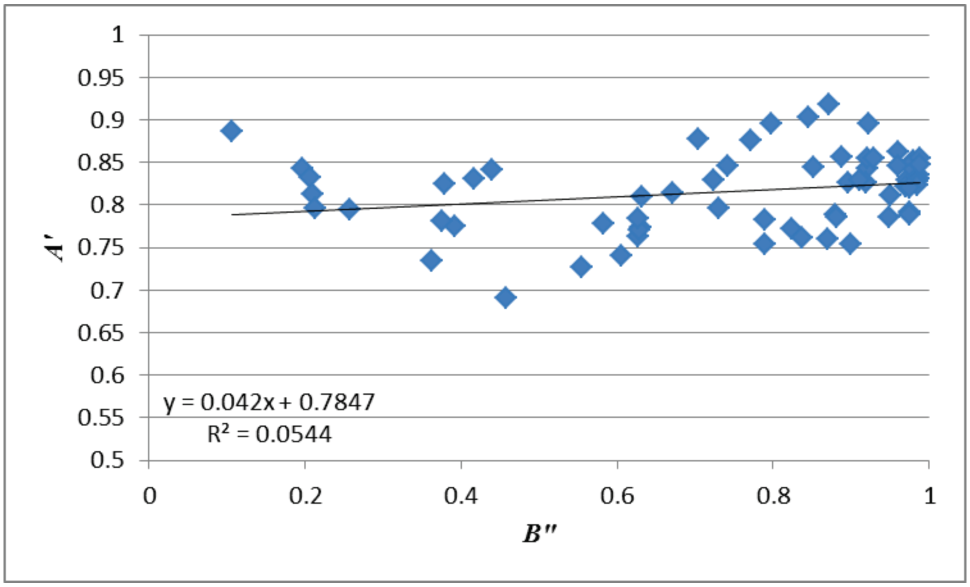
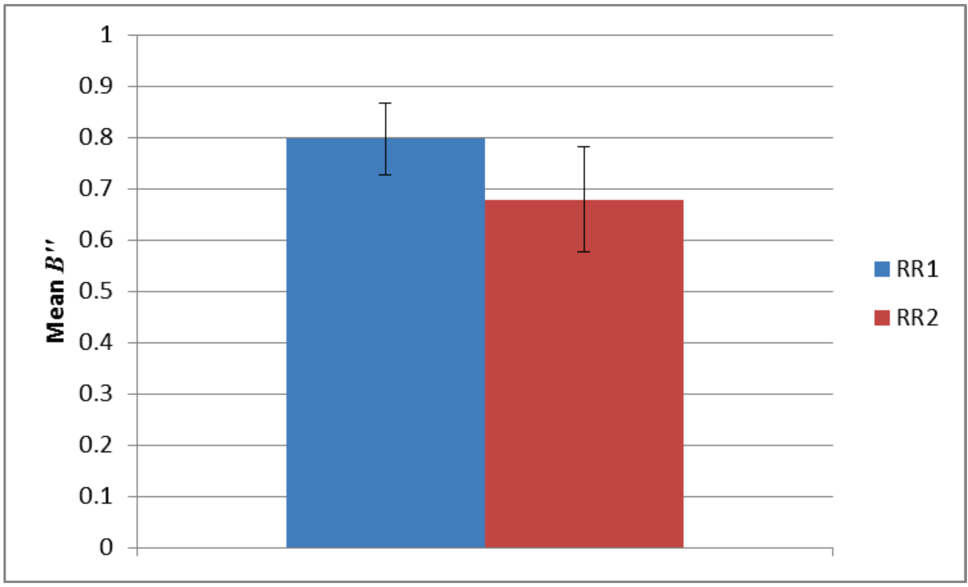


Figure 15. Scatterplot of *A'* and *B''*

As noted in Figure 7, both observation systems (ETG and TI) showed a large bias to say "no." All of the data points for both groups fall below the minor diagonal in Figure 7. Figure 16 is consistent with Figure 7 and it shows that the observation systems on both field tests (RR1 and RR2) have a high bias to say "no", but do not differ statistically with regard to response bias. Other analyses of *B''* are collapsed across the railroads for this reason.



,

Figure 16. Mean *B''* and 95% confidence intervals for RR1 and RR2 across all four observation conditions

Figure 17 shows the mean *B''* values for the ETG and TI systems. The ETG observers show a much higher bias to say "no" than the TI observers, and this difference is statistically reliable. It is also interesting to note that the variability for *B''* is much higher for the TI observers than for the ETG observers.

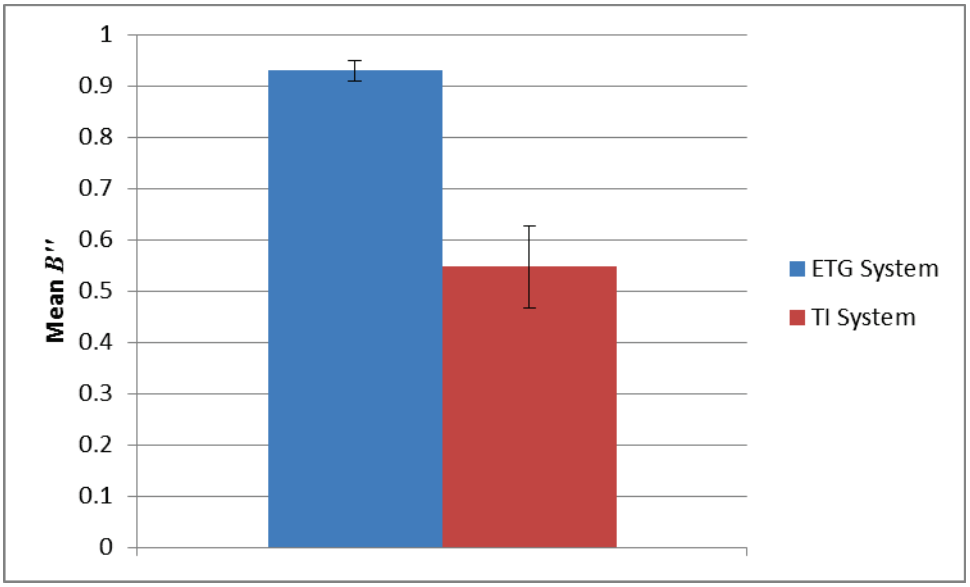


Figure 17. Mean *B''* and 95% confidence intervals for ETGs and TIs across all four observation conditions

Mean *B''* values for the four observation conditions (MR, SC, TGMR, and TGSC) are shown in Figure 18. There is reliably higher bias to say "no" in the MR condition relative to the other three conditions which are not statistically different from each other.

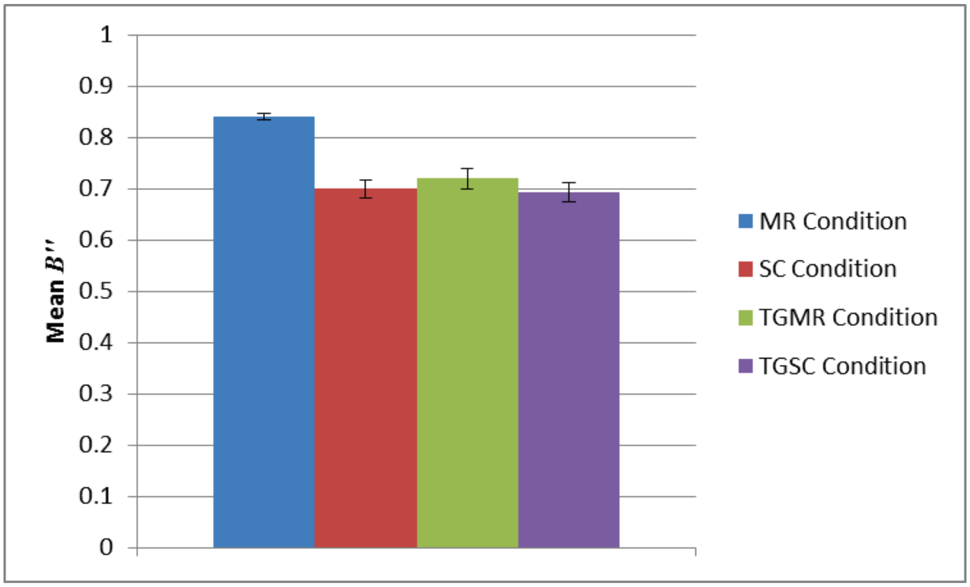


Figure 18. Mean *B''* and 95% confidence intervals for all four observation conditions

The interaction between observation conditions and observation system is shown in Figure 19. It is apparent that bias to say "no" is consistently high for the ETG and TI systems in the MR observation condition and reliably different for all other observation conditions. The ETG observers maintain a consistently higher bias to say “no” across observation conditions, while the TI observers have a consistently lower bias to say “no” in the SC, TGMR, and TGSC conditions.

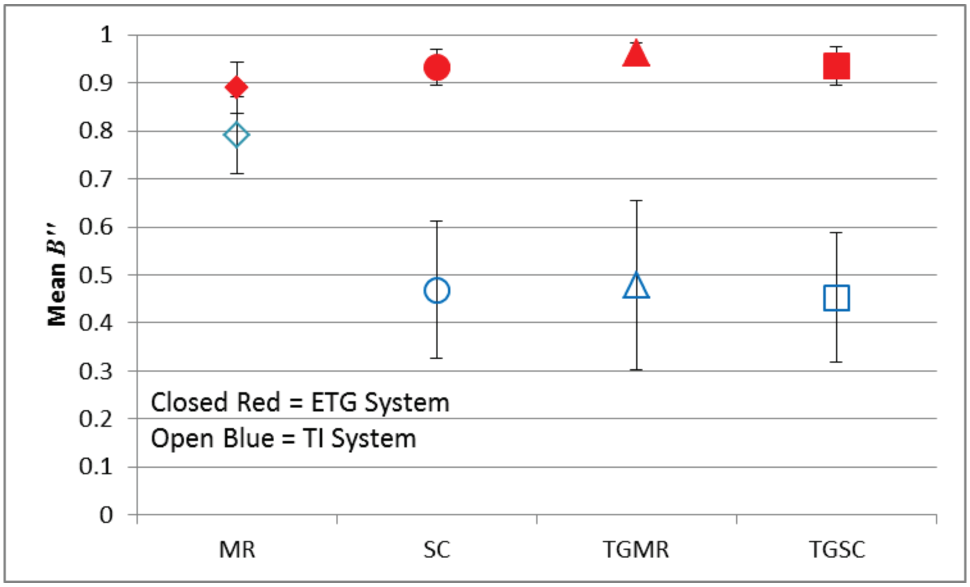


Figure 19. Mean *B’’* and 95% confidence intervals for ETGs and TIs for all four observation conditions

5. Conclusion

Signal Detection Theory (SDT) is a method of analysis that allows the detectability of target events to be analyzed independently from the effects of motivation, attitudinal factors and expectation on the observer's tendency to say "yes" or "no" (response bias) that an event is a target (i.e., a deteriorated track condition). Detectability or sensitivity in SDT can be increased or decreased only by increasing or decreasing the separation of the noise and signal distributions (i.e., changing the signal-to-noise ratio). Response bias is chiefly affected by the values associated with response outcomes as shown in Table 3 and Equation 5, as well as by the prior odds (Equation 4). In the field observations described and analyzed in this report, there was no control over the distributions of signal and noise, the values of the response outcomes, or the prior odds. Nonetheless, there were statistically reliable differences observed in these field tests that are noteworthy and important for the conduct of track inspections.

Preliminary analyses of the ROC data indicated that the data was not normally distributed and required that nonparametric indices of sensitivity (*A'*) and response bias (*B''*) be used. Future research should use existing information about the occurrence of deteriorated track conditions on various classes of track to determine the type of probability distribution that characterizes track defects in a standard length of track. This will enable the use of sensitivity and response bias indices that are specific to the type of distribution and allow the development of quantitative relationships between the SDT parameters and track parameters.

There were no differences observed in sensitivity or response bias between railroads, which indicates that the two railroads were highly similar with regard to the presence of track defects that could be detected under the conditions set by this field test. This allowed data to be combined across railroads for analysis.

With regard to sensitivity or the detectability of deteriorated track conditions, there were no overall differences between the ETG observers and TI observers when averaging among all four observation conditions. This outcome was interesting and slightly unexpected since, in theory, the ETG observation system allows two "looks" at an event before a "yes" or "no" decision is rendered, namely the initial automated TGMS data collection and then the human editing of the exception data. When there are multiple looks in SDT, sensitivity increases (Macmillian and Creelman, 2005). The lack of an overall difference between the ETG and TI systems suggests that human editing of the TGMS data does not constitute a second look at the same information. It is possible that how the TGMS data is used and presented to the human editors could be altered to gain sensitivity for the overall ETG system. Two looks, in theory, could increase sensitivity by 44%.

Detectability did differ between observation conditions, with MR having the least detectability and TGSC having the highest detectability overall. When this was further parsed by the observer system, it was found that the observation condition interacted with the observer system. The TI observers had higher sensitivity for detecting MR conditions than the ETG system, while the opposite was true for the TGSC condition. The two observer systems were statistically equivalent in the other observation conditions.

It is not surprising that the ETG system had a higher sensitivity for detecting a TGSC condition than the TI system. The ETG system (specifically the automated first phase of the system, namely the TGMS) was designed to accurately measure track geometry parameters. At the same time, the ETG system was not designed to detect deteriorated track conditions that have not yet affected track geometry. A significant number of locations that were in the main MR bin but not in the other three sub-bins (SC, TGMR, and TGSC) were related to poor rail surface conditions, several high spikes, and moderately deteriorated tie conditions. Such conditions are generally not readily detectable by the ETG system.

Poor rail surface conditions likely generate high-frequency content that would be digitally filtered out by the ETG system, which is designed to detect wavelengths greater than about four or five feet in length. High spikes may affect lateral stability of the track and would be more readily detectable by automated systems designed specifically to detect poor lateral stability, such as the GRMS. Moderately deteriorated ties could potentially affect track geometry, but this condition would more likely be readily detectable by a TGMS system that was mounted on a full-size railcar that properly loaded the track rather than a TGMS system mounted to a hi-rail vehicle (such as the one used in this study) that provides only a light load on the track and may not properly "seat" the rail.

Furthermore, it is important to point out that none of the bins held a set of "pure" geometry anomalies. Even the TGMR and TGSC sub-bins (despite their names) do not constitute a set of pure geometry conditions, as they include some locations that had a slight geometry anomaly along with another deteriorated condition, such as fouled ballast or high spikes. Achieving an ideal bin of pure geometry anomalies is likely not possible using real-world revenue service track. Instead, a controlled environment is more ideal. The FRA has funded the construction of a track segment at the Transportation Technology Center (TTC) in Pueblo, CO. This track segment is 500 feet in length and allows for the creation of known vertical and lateral geometry deviations through the use of mechanical means, such as shims. Using this sort of controlled environment with known track geometry deviation locations and magnitudes will be a better way to achieve accurate sensitivity parameters for the ETG system, and this approach is currently being investigated.

A condition that was detected by one or more TI observers but not detected by the ETG system can be found in Figure 20. The ballast washout and poor drainage at this location appears to be due to a culvert pipe separation. Eventually, such a condition could worsen and manifest itself in poor track geometry, which may be detected by the ETG system. However, the risk of a derailment due to unstable vertical track support would clearly increase as the condition worsened.

Figure 20. Example of poor drainage condition leading to ballast washout

Overall, the ETG observers had a significantly higher bias to say "no" relative to the TI observers. The ETG points in the ROC plot (Figure 7) also suggested that this observer system was acting as a Neyman-Pearson observer. A Neyman-Pearson observer sets a fixed FA rate and maximizes the Hit rate within that limitation. This is a decision strategy similar to that which is used in statistical hypothesis testing (see Hays, 1963). Here the fixed FA rate or probability corresponds to the setting of an α level for the probability of a Type I error (by convention 0.05 or 0.01). The mean FA probability for the ETG observers was 0.007±0.002. For the TI observers the mean FA probability was 0.08±0.02. So, there is an order of magnitude difference for both the mean FA and the standard error.

It does not appear that the TI observers are acting as Neyman-Pearson observers, and the question is "Why are these two groups using different decision strategies?" It is possible that each group has different norms for acceptable performance that is communicated formally or informally. For instance, the human editors involved in the ETG system may have been trained to avoid FAs and reprimanded for excessive FAs. TI observers may have been trained to report any possible defect regardless of its status upon verification. Also, as was stated earlier in the report, the TI system of inspection is typically a two-stage process where the TI observer initially inspects the track from a hi-rail vehicle and then leaves the hi-rail vehicle to conduct a more thorough ground inspection when they spot a potential anomalous track condition from within the hi-railer. This second stage may lead the TI observer to reconsider their initial suspicion that the location was deteriorated, and therefore, they may choose not to include the location in a final list of deteriorated track locations. Due to time constraints and overall logistical considerations, TI observers participating in the study were asked to stay in the hi-rail vehicle. Therefore, the first phase of the TI system, namely the initial detection of a potential deteriorated track condition from the hi-rail vehicle, was the only phase tested. If the second phase of the typical TI process was included in the study, it likely would have led to fewer FAs for the TI system since TI observers may have chosen to discard certain locations upon conducting a more thorough ground investigation. Reducing the number of FAs may have slightly increased the sensitivity of the TI system overall.

Further research is needed to determine what factors are influencing decision making in the TI system and ETG system so that the differences in response bias can be understood and used to improve the track inspection process. A focus on reducing FAs restricts the number of Hits because Hits and FAs co-vary as seen in Figure 3. Because Hits and Misses come from the same probability distribution, a decrease in Hits also results in an increase in Misses, which can result in accidents. So setting a very low FA rate may save money in the short run but could lead to costlier accidents in the long run. The mean Miss probability was 0.66 and 0.57 for ETG observers and TI observers, respectively. Not all missed track defects will result in an accident, but a high rate of Misses raises the probability that an accident will occur. These values can be changed by changing the payoff matrix which is causing both observer systems to be biased to respond "no."

Response bias to say "no" is also significantly higher in the MR condition relative to all the other observation conditions. This may be due to the fact that the MR bin is the main bin (Figure 1) and includes a significant number of conditions that were classified as a maintenance concern but were not considered a safety critical condition. Conditions that are a maintenance concern but not a safety critical condition can be somewhat subjective and what one track professional considers a maintenance concern (but not a safety concern) another might not consider a maintenance concern. On the other hand, there is likely more agreement on what constitutes a safety critical condition.

As with sensitivity, response bias also has an interaction between observation conditions and observer systems. TI observers have reliably less response bias to say "no" than ETG observers in all observation conditions except the MR condition. ETG observers have a consistent bias to say "no" across all observation conditions. This may be another manifestation of the effects of training or other processes that cause this group to avoid FAs. Also, there is a high likelihood that the high Miss rate for the ETG system with regard to MR conditions is due to the criterion levels set by the first phase of the ETG system, namely the automated data collection via a TGMS. This system uses track geometry levels or thresholds (gage, crosslevel, alinement, and profile) put forth in the TSS. These levels are meant to be *minimum safety standards*. Therefore, they are more in tune with the TGSC condition being considered in this study.

Lowering the threshold levels for the various geometry parameters in the TGMS software would lead to a greater number of exceptions for the human editors (the second phase of the ETG system) to review and consider. For example, Section 213.63 – "Track Surface" of the TSS puts forth a value of 2 inches for the Class 4 track surface's (also known as track profile) 62-foot mid-chord offset. However, if this level was lowered to 1.5 inches, for example, in the TGMS software, then more exception locations would be output by the TGMS system for review by the human editors. In addition to profile, reduction of such threshold values could be applied to other track geometry parameters, such as gage, crosslevel, and alinement. Inevitably this would result in a lower Miss rate but also a higher FA rate, and it is likely that the overall sensitivity of the ETG system would not change significantly.

This report has focused on the ability of the TI and ETG systems to detect deteriorated track conditions. The high probability of Misses in both systems may be viewed by some as an example of "human error" or a "misjudgment" on the part of the TIs and the editors of the track geometry data (the human component of the ETG system). This view, implicitly or explicitly, wrongly blames the operators and inspectors for decision outcomes that may largely be determined by organizational policy and management. Misses are an inevitable consequence of human behavior when a binary decision ("yes, there is a deteriorated track condition"; "no, there is not a deteriorated track condition") is made under conditions of uncertainty.

A high probability of Misses is inevitable if the goal of organizational policy and management is to keep the probability of FAs low. The probability of FAs in this report was very low for both the TI and ETG systems. The consequence of a low FA rate is a low Hit rate because Hits and FAs co-vary (Figure 3). Since the probability of Hits and Misses sum to one, a low FA rate results in a high Miss rate. FAs are costly because they disrupt operations and cause delays needlessly, so it is understandable that management's policy might be to keep FAs low. This might be accomplished through training, supervisor feedback, or other positive and negative incentives that constitute the value of an FA (V(FA) in Table 3). However, if the policy only focuses on FAs, it ignores the value of a Miss (V(Miss)) which includes the cost of accidents. In order to be optimal, policy must recognize that there is a trade-off between FAs and Misses (see Equation 5).

We do not know what organizational policies (explicit or implicit) are in effect that may have caused the high level of bias to say "no" in this study and resulted in a high probability of Misses. That would be the study for a future project. We do know that prior odds and the payoff matrix determine response bias. These factors may have set up the TIs and editors of the track geometry data to behave in predictable ways. It is wrong to consider their behavior in this study as "human error" or "misjudgment." Rather, it may be predictable human behavior.

In addition, it is important to realize that a bias to say "no," which results in a high Miss rate, may be acceptable if the inspection frequency is high enough and/or the degradation rate of the deteriorated track condition is low enough. A risk assessment model based on Markov chain theory or Monte Carlo methods would take the Miss rate as well as the inspection frequency and degradation rates in order to optimize inspection strategies. In other words, the Miss rate should not be considered on its own, as it is inherently linked to inspection frequency and degradation rates.

Future studies may be directed towards investigating the sources of bias in both the ETG and TI system. As was stated previously with regard to the ETG system, none of the bins constituted a set of "pure" geometry anomalies. Achieving a set of pure geometry anomalies is more realistic in a controlled environment, such as the test track at TTC that was described earlier in this section. Further studies are planned to test the sensitivity and bias of the ETG system using this track, which likely will result in more dependable values for each parameter. In addition, future studies may investigate the effect of the second phase of the TI system, namely allowing TI observers to exit the hi-rail vehicle, on the overall sensitivity of the TI system. Also, the effect of double track or triple track inspection on TI system sensitivity should be investigated. Visual search time for a specific item (e.g., deteriorated track conditions) increases with the number of items available to search (multiple tracks) and affects detection (Al-Nazer, Raslear, Patrick, Gertler, Choros, Gordon, and Marquis, 2011). Finally, future studies may be directed at quantifying the sensitivity of newer technologies that are being adopted by the industry, such as the GRMS and joint bar inspection systems.

6. References

Al-Nazer, L., Raslear, T. G., Patrick, C., Gertler, J., Choros, J., Gordon, J., & Marquis, B. (2011). *Track inspection time study* (Report No. DOT/FRA/ORD-11/15). Washington, DC: U.S. Department of Transportation. Available at https://www.fra.dot.gov/eLib/details/L01299#p1_z5_gD_ktime%20study.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

Daniel, W. W. (1978). *Biostatistics: A foundation for analysis in the health sciences* (2nd ed.). New York: Wiley.

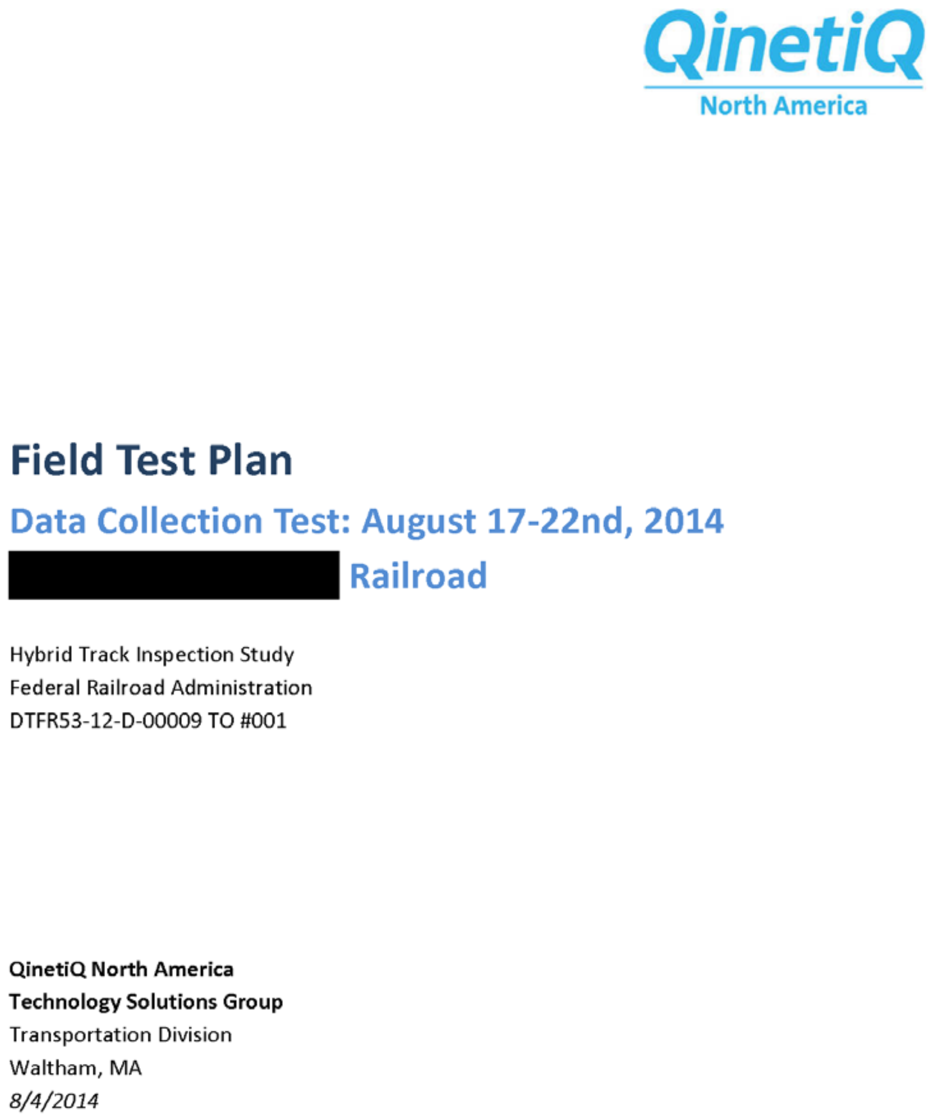
Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Green, D. M. & Swets, J. W. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hays, W. L. (1963). *Statistics for psychologists*. New York: Holt, Rinehart & Winston.

Macmillan, N. A. & Creelman, D. C. (2005). *Detection theory: A user's guide* (2nd ed.). Mahway, New Jersey: Lawrence Erlbaum.

Appendix A.
RR1 Test Plan



QinetiQ
North America

Field Test Plan

Data Collection Test: August 17-22nd, 2014

[redacted] Railroad

Hybrid Track Inspection Study
Federal Railroad Administration
DTFR53-12-D-00009 TO #001

QinetiQ North America
Technology Solutions Group
Transportation Division
Waltham, MA
8/4/2014

Introduction

QinetiQ North America (QNA) is executing an FRA-sponsored project to characterize the accuracy and sensitivity of various systems, both human and automated, that inspect track. For certain conditions, humans excel at track inspection; whereas, other conditions are more readily detected by automated systems. Determining the probability of detection (POD) of a deteriorated track condition for each of these methods of inspection (human and automated) will allow for the formulation of an optimal hybrid inspection strategy in order to minimize risk of a derailment.

The human and automated inspection systems share the common goal of accurately identifying all points along the track that have a deteriorated condition or have the potential to deteriorate in the near future (hereinafter "points of interest"), but the process by which they perform this task differs. Human track inspectors inspect the track by driving a track inspection vehicle over the track. As they drive, they visually inspect the track looking for points of interest. The automated track inspection vehicle drives along the track as well and scans the track using a proprietary inertial and laser-based system known as a track geometry measurement system (TGMS). When an anomaly is found, the system generates an "exception." A trained individual then filters these exceptions manually (based on other data presented graphically) to determine if the point of interest is valid or if the point of interest was inaccurately reported.

QNA will utilize signal detection theory (SDT) to compare the responses from the human and automated inspections systems to the ground truth list of points of interest. SDT is an analysis tool that will determine the sensitivities of both systems (human and machine) for detecting different types of points of interest (POIs). SDT can also determine if differences in performance are due to response bias (i.e., system has a high detection rate but at the cost of a high false alarm rate. This type of statistical analysis will not only provide a generalized form of measurement for both systems, but it also allows us to discern far more information about the systems we are investigating than simply looking at POIs correctly and incorrectly identified. Using SDT to analyze both human and automated inspection methods will give us a better idea as to the strengths and weaknesses that each system possesses. In identifying the vulnerabilities that either system may have, we can develop an optimized hybrid inspection system for a safer, more efficient inspection of track. These ground truth points of interest that will be used for SDT analysis will be obtained by an independent consultant who will inspect the track post-experiment by taking detailed track measurements as defined by Federal Track Safety Standards (FTSS), 49 CFR Part 213. The independent consultant will provide his findings to QNA in the form of a report.

Test Location

Data collection will occur at the [redacted] Railroad between [redacted] and [redacted], Virgina. This section of track will be approximately 10-15 miles in length.

Types of Points of Interest

A Point of Interest (POI) is any location that would prompt documentation as an FRA safety violation, possible maintenance concern, or necessitate further ground inspection. In a typical inspection you would normally document safety violations as a ticket, etc. For the purposes of this study this will not be necessary. It is not within the scope of this study for inspectors to be submitting violation reports or tickets.

You will be a passenger in the hi-railer and can direct the driver to change speed, stop, or reverse when necessary to perform a thorough visual track inspection. However, the hi-railer cannot exceed track speed limits or safe operating speed limits for hi-rail vehicles. When you perceive that the hi-railer is within the zone of the POI, we ask that you press the button and verbally describe the POI. A POI can be of two different types: a single location on track (e.g broken tie), or a length of track (e.g low ballast.) For the former, push the button when the hi-railer is at the location of the POI. For the latter, push the button at the beginning of the length of track where the POI exists. GPS location will be recorded with each button press to indicate the location of the POI.

Test Participants

There are two categories of test participants; track inspectors and exception editors. The track inspectors will conduct the human inspection of the track while the exception editors will review the exception reports from the TGMS to identify and filter out false hits.

Track Inspectors

Test participants will be recruited from FRA track inspectors. Because the study will be conducted as part of their normal working hours, no special compensation for participation will be provided. Travel for the participants will be borne directly by FRA. Four participants per location will be arranged. All FRA track inspectors are assumed to be experts with several years of experience.

Exception Editors

Test participants will be provided by [redacted]. Because the study will be conducted as part of their normal working hours, no special compensation for participation will be provided. The filtering process will be conducted post data collection and off-site at an [redacted] location, and as such, travel costs are not anticipated. All [redacted] exception editors are assumed to be experts with several years of experience. Four editors per test will be required.

Test Equipment

The following equipment will be required for the study:

1. Hi-railer for human inspectors – supplied by [redacted]
2. Track Geometry Measurement System and vehicle – supplied by [redacted]
3. Tools for ground inspection verification, specifically a string lining kit, gauge level board, and a 18-inch straight edge – supplied by [redacted]
4. Eye-tracking system – supplied by QNA

5. Push Button Data Acquisition system– supplied by QNA

Participant Response Data Collection

Track Inspectors

Human inspector response data will be collected utilizing a button press/auditory comment data acquisition system built on LabView architecture. This application will be designed to minimize the cognitive loading on the inspector and automate many of the data collection tasks. The participants will record their responses (points of interest) using a handheld button "wand" and describing whatever type of POI they see. Our data acquisition system is designed to capture audio and GPS data upon each button press. While participants are imputing their audio responses, QNA will be secondarily recording and categorizing their responses to streamline post collection data analysis procedures and efficiency.

In addition, a no contact eye-tracking system will be utilized to determine where the inspector is looking to help confirm if they have correctly identified a point of interest. Finally, an internal microphone and scene camera will record the inspector's spoken responses in identifying other types of points of interest not included in this study. GPS location and timestamp will be used to correlate data between the human inspector and the TGMS.

Exception Filterers

The response of the exception filterers will be collected using the iPad data acquisition application (DAA). An exception report will be pre-fed into the iPad DAA and the editors will respond to each exception on the same 2-point scale detailed previously. To inform their decisions regarding each exception, the editors will have access to the raw TGMS data. Both one class drops and two class drops will be analyzed.

Schedule

The second data collection test will occur at [redacted] Railroad between August 17, and August 22, 2014. Below is the anticipated schedule of events.

Day 1

- <u>9:00am:</u> [redacted] will travel with track master to track and begin their track inspections.
- <u>9:15am or 10:45am:</u> [redacted] will perform their first two track inspections.
- <u>11:00am – 12:00pm</u>: Break for early lunch.
- <u>12:15pm or 1:45pm:</u> [redacted] will perform their second two track inspections.
- Throughout this day QNA will be preparing the poi/exceptions list for the final ground verification.
- Thresholds for various geometry parameters will be adjusted in order to obtain a reasonable list for ground verification.

Day 2

- Task One: First Field Validation
  - QNA and Consultant will arrive at Test site location in the morning.
  - We will meet roadmaster/railroad representative and travel with consultant to track location.
  - First track validation will take place.
  - When first ground verification is done, QNA will analyze this data and use it to prepare specifics of the data acquisition system.
- Task Two: Debrief Inspectors
  - FRA inspectors will arrive in late afternoon.
  - At this time all forms will be filled out and any questions of any kind the inspectors may have will be answered.
  - At this time we will also conduct safety briefing.

Day 3

- 9:15am: Meet first participant at either participant hotel *(9:15am)* or beginning of test site conduct safety briefing, begin system calibration and provide participant test instructions.
- 10:30am – 12:00pm: Conduct first test.
- 12:00pm: Release test participant.
- 12:00pm – 1:00pm: Break for lunch.
- 2:30pm – 4:00pm: Conduct second test.
- 4:00pm: Release test participant.
- 4:15pm: QNA remove equipment from hi-railer.
- Evening: QNA to review first day's data with track consultant. FRA track inspector representative will be available for consultation if needed.

Day 4

- 9:00am: FRA inspectors and QNA meet at hotel parking lot.
- 9:15am or 10:00am: Meet first participant at either participant hotel *(9:15am)* or beginning of test site conduct safety briefing, begin system calibration and provide participant test instructions.
- 10:30am – 12:00pm: Conduct first test.
- 12:00pm: Release test participant.
- 12:00pm – 1:00pm: Break for lunch.
- 1:15pm or 2:00pm: Meet second participant at either participant hotel *(1:15pm)* or beginning of test site), conduct safety briefing, begin system calibration and provide participant test instructions.
- 2:30pm – 4:00pm: Conduct second test.

- 4:00pm: Release test participant.
- 4:15pm: QNA remove equipment from [redacted] hi-railer.
- Evening: QNA to review first and second day's data with track consultant. FRA track inspector representative will be available for consultation if needed.

Day 5

- 9:00am: QNA, track consultant, and [redacted] meet at test site for safety briefing
- 9:15am – 6:00pm: Conduct ground truth verification

Day 6

- 9:00am: QNA, track consultant, and [redacted] meet at test site for safety briefing
- 9:15am – 6:00pm: Conduct ground truth verification if needed

See below:

Sunday, August 17	Monday, August 18	Tuesday, August 19	Wednesday, August 20	Thursday, August 21	Friday, August 22
[redacted] runs	1st Ground Verification	FRA run 1	FRA run 2	2nd Verification	2nd Verification
Data collection/ analysis/consolidation		Data collection	Data collection		

Test Procedure

Data Collection at Track

1. Upon arrival at track, each participant will be asked to read the consent form. QNA will stress that participation is voluntary and can be ended at any time at the participant's discretion. QNA will explain that all participant specific data collected will only be utilized for study purposes and will not be shared with FRA except in aggregate or anonymously. QNA will stress that their performance in this study will affect their job in no way. QNA will ask each participant if they have any specific safety or privacy concerns. Once all questions are resolved and the participant is satisfied, he/she will be asked to sign the consent form.
2. After the participant has read and signed the consent form, they will be provided test instructions. Based on the results of the consultant's initial track inspection, FRA inspectors will be asked to inspect the track to a specific FRA class standard according to 49CFR213.13 parts A-F and specific class of maintenance. They will be asked to record all points of interest they see by pressing the button and verbally describing what they see or feel. QNA will give each participant instructions on operating the button DAQ system. Lastly, QNA will instruct the participant that they may travel at a variable speed in the hi-railer, and if they need they may stop the vehicle and backup. They may travel as slowly as they need to do a thorough and fastidious inspection,

but we encourage maintaining a minimum speed of 5 miles per hour. Exiting the vehicle is not permitted, except in the rare circumstance where a safety critical defect is present.

3. During FRA inspection, a high rail truck will be driven in which John and Duncan will be acquiring data and providing any additional instruction or providing answers to any questions inspectors may have.
4. When the human inspection portion of the study is finished, the [redacted] portion of the study will begin. The TGMS vehicle will begin track inspection and data collection. It will travel at the maximum allowed track speed.
5. The second ground verification will begin on the last two days of the study. All those involved in the study will be welcomed to participate in aiding the consultant in his validation of POI found throughout the previous days of data collection.

Data Collection Off-Site with Exception Editors

1. Shortly after the on-site data collection has been completed, QNA will travel to [redacted]'s site to collect TGMS exception report filtering data.

2. QNA will pre-load one of the collected exception reports into the iPad data collection application.

3. Each participant will then be asked to read the consent form. As before, QNA will stress that participation is voluntary and can be ended at any time at the participant's discretion. QNA will explain that all participant specific data collected will only utilized for study purposes and will not be shared with FRA or their employer except in aggregate or anonymously. QNA will stress that their performance in this study will affect their job in no way. QNA will ask each participant if they have any specific safety or privacy concerns. Once all questions are resolved and the participant is satisfied, he/she will be asked to sign the consent form.

4. QNA will give each participant instructions on how to operate the iPad application and how to record responses.

5. The editor will then filter the exception reports as they would normally. They will be presented with a playback of the TGMS data feed/ strip charts and asked to filter the exceptions using their normal decision making process to class three standards. They will be asked to record their responses using the iPad data collection application instead of their normal software.

6. At the end data collection, the editor will be released and steps 1-5 will repeated for the next three exception reports, utilizing a different editor for each report.

Safety

Safety on track will be paramount to all other considerations. The following measures will be taken to ensure site safety.

- All participants and individuals will be required to wear safety shoes, safety vests, hard hats and safety glasses while on track or in the vicinity of track.
- [redacted] personnel will final have authority over the operation of their vehicles on and off track to ensure the safety of their personnel and vehicles.
- A railroad representative will be on site to monitor safety and to communicate with the railroad dispatcher. Track warrant for the complete length of test track will be obtained during testing operations.
- The railroad representative will conduct a safety briefing at the beginning of each data collection test.
- If two high railers are used for any portion of this study, communication will be maintained between both hi-railers via 2-way radios. If the lead hi-railer makes an emergency stop while on track, they are to inform the following vehicle immediately.
- Both [redacted] hi-railers will shunt the track to activate the gates for gated crossings. However there are can be numerous highway grade crossings without gates. At these crossings, the hi-railers will slow down to a safe speed (as determined by [redacted] personnel) or come to a full stop to clear the crossing before proceeding.

Appendix B
RR2 Test Plan

QinetiQ
North America

Field Test Plan

Data Collection Test: November 10th-15th, 2014
[redacted] Railroad

Hybrid Track Inspection Study
Federal Railroad Administration
DTFR53-12-D-00009 TO #001

QinetiQ North America
Transportation Division
Waltham, MA
11/6/2014

Introduction

QinetiQ North America (QNA) is executing an FRA-sponsored project to characterize the accuracy and sensitivity of various systems, both human and automated, that inspect track. For certain conditions, humans excel at track inspection; whereas, other conditions are more readily detected by automated systems. Determining the probability of detection (POD) of a deteriorated track condition for each of these methods of inspection (human and automated) will allow for the formulation of an optimal hybrid inspection strategy in order to minimize risk of a derailment.

The human and automated inspection systems share the common goal of accurately identifying all points along the track that have a deteriorated condition or have the potential to deteriorate in the near future (hereinafter "points of interest"), but the process by which they perform this task differs. Human track inspectors inspect the track by driving a track inspection vehicle over the track. As they drive, they visually inspect the track looking for points of interest. The automated track inspection vehicle drives along the track as well and scans the track using a proprietary inertial and laser-based system known as a track geometry measurement system (TGMS). When an anomaly is found, the system generates an "exception." A trained individual then filters these exceptions manually (based on other data presented graphically) to determine if the point of interest is valid or if the point of interest was inaccurately reported.

QNA will utilize signal detection theory (SDT) to compare the responses from the human and automated inspections systems to the ground truth list of points of interest. SDT is an analysis tool that will determine the sensitivities of both systems (human and machine) for detecting different types of points of interest (POIs). SDT can also determine if differences in performance are due to response bias (i.e., system has a high detection rate but at the cost of a high false alarm rate. This type of statistical analysis will not only provide a generalized form of measurement for both systems, but it also allows us to discern far more information about the systems we are investigating than simply looking at POIs correctly and incorrectly identified. Using SDT to analyze both human and automated inspection methods will give us a better idea as to the strengths and weaknesses that each system possesses. In identifying the vulnerabilities that either system may have, we can develop an optimized hybrid inspection system for a safer, more efficient inspection of track. These ground truth points of interest that will be used for SDT analysis will be obtained by an independent consultant who will inspect the track post-experiment by taking detailed track measurements as defined by Federal Track Safety Standards (FTSS), 49 CFR Part 213. The independent consultant will provide his findings to QNA in the form of a report.

Test Location

Data collection will occur at the [redacted] Railroad in [redacted], Texas. The start and end points of the tests are to be determined. The section of track will be approximately 10-15 miles in length.

Types of Points of Interest

A Point of Interest (POI) is any location that would prompt documentation as an FRA safety violation, possible maintenance concern, or necessitate further ground inspection. In a typical inspection you would normally document safety violations as a ticket, etc. For the purposes of this study this will not be necessary. It is not within the scope of this study for inspectors to be submitting violation reports or tickets.

You will be a passenger in the hi-railer and can direct the driver to change speed, stop, or reverse when necessary to perform a thorough visual track inspection. However, the hi-railer cannot exceed track speed limits or safe operating speed limits for hi-rail vehicles. When you perceive that the hi-railer is within the zone of the POI, we ask that you press the button and verbally describe the POI. A POI can be of two different types: a single location on track (e.g broken tie), or a length of track (e.g low ballast.) For the former, push the button when the hi-railer is at the location of the POI. For the latter, push the button at the beginning of the length of track where the POI exists. GPS location will be recorded with each button press to indicate the location of the POI.

Test Participants

There are two categories of test participants; track inspectors and exception editors. The track inspectors will conduct the human inspection of the track while the exception editors will review the exception reports from the TGMS to identify and filter out false hits.

Track Inspectors

Test participants will be recruited from FRA track inspectors. Because the study will be conducted as part of their normal working hours, no special compensation for participation will be provided. Travel for the participants will be borne directly by FRA. Four participants per location will be arranged. All FRA track inspectors are assumed to be experts with several years of experience.

Exception Editors

Test participants will be provided by [redacted]. Because the study will be conducted as part of their normal working hours, no special compensation for participation will be provided. The filtering process will be conducted post data collection and off-site at an [redacted] location, and as such, travel costs are not anticipated. All [redacted] exception editors are assumed to be experts with several years of experience. Four editors per test will be required.

Test Equipment

The following equipment will be required for the study:

1. Hi-railer for human inspectors – supplied by [redacted]
2. Track Geometry Measurement System and vehicle – supplied by [redacted]
3. Tools for ground inspection verification, specifically a string lining kit, gauge level board, and a 18-inch straight edge – supplied by [redacted]
4. Push Button Data Acquisition system– supplied by QNA

Participant Response Data Collection

Track Inspectors

Human inspector response data will be collected utilizing a button press/auditory comment data acquisition system built on LabView architecture. This application will be designed to minimize the cognitive loading on the inspector and automate many of the data collection tasks. The participants will record their responses (points of interest) using a handheld button "wand" and describing whatever type of POI they see. Our data acquisition system is designed to capture audio and GPS data upon each button press. While participants are imputing their audio responses, QNA will be secondarily recording and categorizing their responses to streamline post collection data analysis procedures and efficiency. An internal microphone will record the inspector's spoken responses in identifying points of interest in this study; this will facilitate subsequent data analysis. GPS location and timestamp will be used to correlate data between the human inspector and the TGMS.

Exception Filterers

The response of the exception filterers will be collected using the iPad data acquisition application (DAA). An exception report will be pre-fed into the iPad DAA and the editors will respond to each exception on the same 2-point scale detailed previously. To inform their decisions regarding each exception, the editors will have access to the raw TGMS data. Both one class drops and two class drops will be analyzed.

Schedule

The second data collection test will occur at [redacted] Railroad between November 10th, and November 15th, 2014. Below is the anticipated schedule of events.

Day 1

- 7:30am: [redacted] will travel with track master to track and begin their track inspections.
- 7:45am or 9:45am: [redacted] will perform their first two track inspections.
- 10:00am – 11:00pm: Break for early lunch.
- 12:00pm or 1:45pm: [redacted] will perform their second two track inspections.
- Throughout this day QNA will be preparing the poi/exceptions list for the final ground verification.
- Thresholds for various geometry parameters will be adjusted in order to obtain a reasonable list for ground verification.

Day 2

- Task One: First Field Validation
    - QNA and Consultant will arrive at Test site location in the morning.
    - We will meet roadmaster/railroad representative and travel with consultant to track location.
    - First track validation will take place.
    - When first ground verification is done, QNA will analyze this data and use it to prepare specifics of the data acquisition system.

Day 3

- <u>8:00am:</u> Meet first participant [redacted] to conduct safety briefing, begin system calibration and provide participant test instructions.
- <u>8:45am – 11:00am:</u> Conduct first test.
- <u>11:00pm:</u> Release test participant.
- <u>11:00pm – 12:00pm</u>: Break for lunch.
- <u>12:30pm – 12:45pm</u>: Meet second participant [redacted] to conduct safety briefing, begin system calibration and provide participant test instructions.
- <u>12:45pm – 3:00pm:</u> Conduct second test.
- <u>3:00pm:</u> Release test participant.
- <u>3:15pm:</u> QNA remove equipment from hi-railer.
- <u>Evening:</u> QNA to review first day's data with track consultant. FRA track inspector representative will be available for consultation if needed.

Day 4

- <u>8:00am:</u> Meet first participant [redacted] to conduct safety briefing, begin system calibration and provide participant test instructions.
- <u>8:45am – 11:00am:</u> Conduct first test.
- <u>11:00pm:</u> Release test participant.
- <u>11:00pm – 12:00pm</u>: Break for lunch.
- <u>12:30pm – 12:45pm</u>: Meet second participant [redacted] to conduct safety briefing, begin system calibration and provide participant test instructions.
- <u>12:45pm – 3:00pm:</u> Conduct second test.
- <u>3:00pm:</u> Release test participant.
- <u>3:15pm:</u> QNA remove equipment from hi-railer.
- <u>Evening:</u> QNA to review first day's data with track consultant. FRA track inspector representative will be available for consultation if needed.

Day 5

- <u>8:00am:</u> QNA, track consultant, and [redacted] meet at test site for safety briefing
- <u>8:15am – 6:00pm:</u> Conduct ground truth verification

Day 6

- <u>8:00am:</u> QNA, track consultant, and [redacted] meet at test site for safety briefing
- <u>8:15am – 6:00pm:</u> Conduct ground truth verification if needed

See below:

Monday, November 10th	Tuesday, November 11th	Wednesday, November 12th	Thursday, November 13th	Friday, November 14th	Saturday, November 15th
[redacted] runs	1st Ground Verification	FRA run 1	FRA run 2	2nd Ground Verification	2nd Ground Verification
Data collection/ analysis/consolidation		Data collection	Data collection		

Test Procedure

Human Inspector Data Collection

1. Upon arrival at track, each participant will be asked to read the consent form. QNA will stress that participation is voluntary and can be ended at any time at the participant's discretion. QNA will explain that all participant specific data collected will only be utilized for study purposes and will not be shared with FRA except in aggregate or anonymously. QNA will stress that their performance in this study will not affect their job in any way. QNA will ask each participant if they have any specific safety or privacy concerns. Once all questions are resolved and the participant is satisfied, he/she will be asked to sign the consent form.
2. After the participant has read and signed the consent form, they will be provided test instructions. Based on the results of the consultant's initial track inspection, FRA inspectors will be asked to inspect the track to a specific FRA class standard according to 49CFR213.13 parts A-F and specific class of maintenance. They will be asked to record all points of interest they see by pressing a button and verbally describing what they see or feel. QNA will give each participant instructions on operating the button on the data collection system. Lastly, QNA will instruct the participant that they may travel at a variable speed in the hi-railer, and if they need to, they may stop the vehicle and backup. They may travel as slowly as they need to do a thorough inspection, but we encourage maintaining a minimum speed of 5 miles per hour. Exiting the vehicle is not permitted, except to inspect switches, if necessary, and in the rare circumstance where a safety critical defect is present.
3. During FRA inspection, a high rail truck will be driven in which the Principal Investigator and Data Collection Analyst will be acquiring data and providing any additional instruction or providing answers to any questions inspectors may have.
4. A preliminary ground inspection will be performed to assist in data analysis. This will happen prior to the human track inspection data collection. Inspection to specific class track will be determined based on the preliminary ground inspection. Ground verification will be conducted

on the last two days of the study. All those involved in the study are welcome to participate in aiding the consultant in his validation of POIs found throughout the previous days of data collection.

Data Collection Off-Site with Exception Editors

1. After the on-site data collection has been completed, QNA will travel to [redacted]'s site to collect TGMS exception report filtering data.

2. QNA will pre-load one of the collected exception reports into the iPad data collection application.

3. Each participant will then be asked to read the consent form. As before, QNA will stress that participation is voluntary and can be ended at any time at the participant's discretion. QNA will explain that all participant specific data collected will only utilized for study purposes and will not be shared with FRA or their employer except in aggregate or anonymously. QNA will stress that their performance in this study will affect their job in no way. QNA will ask each participant if they have any specific safety or privacy concerns. Once all questions are resolved and the participant is satisfied, he/she will be asked to sign the consent form.

4. QNA will give each participant instructions on how to operate the iPad application and how to record responses.

5. The editor will then filter the exception reports as they would normally. They will be presented with a playback of the TGMS data feed/ strip charts and asked to filter the exceptions using their normal decision making process to class three standards. They will be asked to record their responses using the iPad data collection application instead of their normal software.

6. At the end data collection, the editor will be released and steps 1-5 will repeated for the next three exception reports, utilizing a different editor for each report.

Safety

Safety on track will be paramount to all other considerations. The following measures will be taken to ensure site safety.

- All participants and individuals will be required to wear safety shoes, safety vests, hard hats and safety glasses while on track or in the vicinity of track.
- [redacted] personnel will final have authority over the operation of their vehicles on and off track to ensure the safety of their personnel and vehicles.

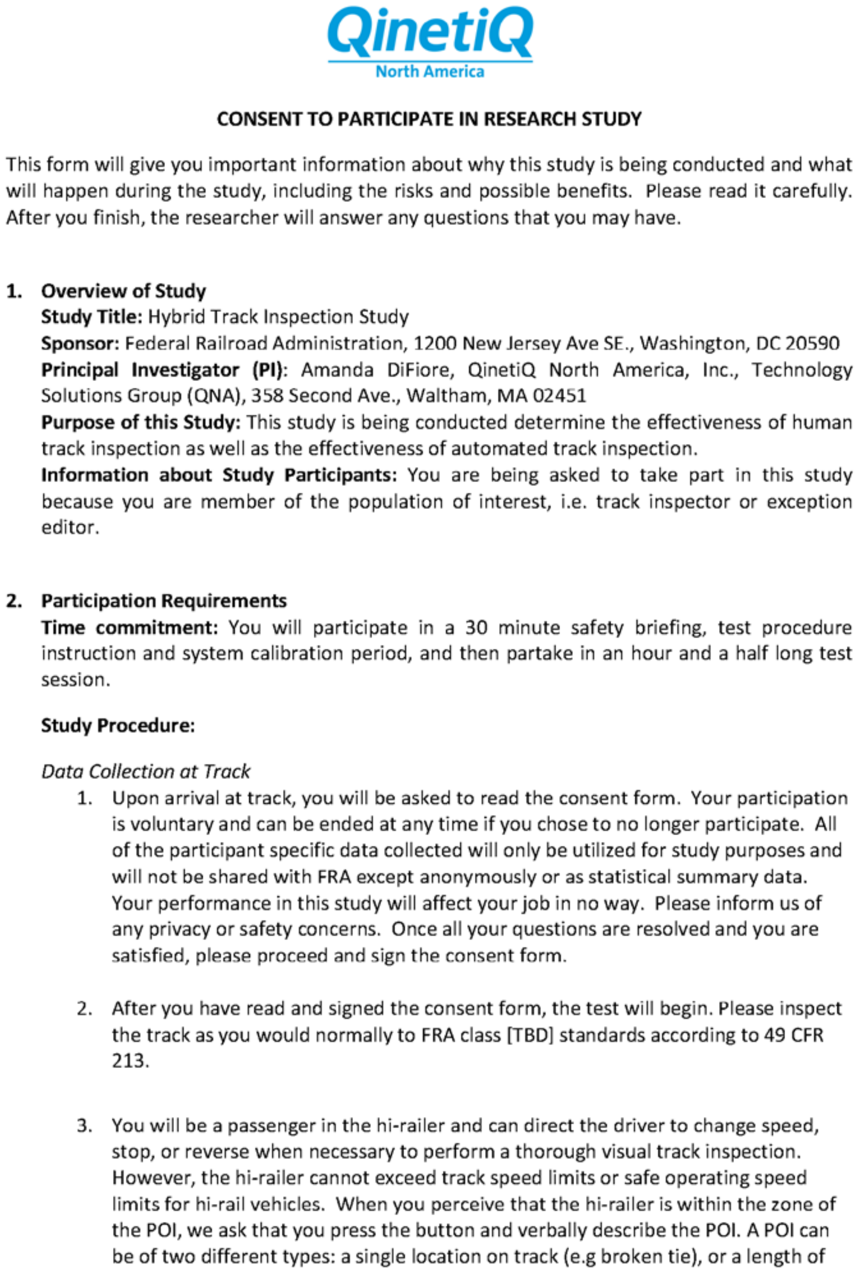
- A railroad representative will be on site to monitor safety and to communicate with the railroad dispatcher. Track warrant for the complete length of test track will be obtained during testing operations.
- The railroad representative will conduct a safety briefing at the beginning of each data collection test.
- If two high railers are used for any portion of this study, communication will be maintained between both hi-railers via 2-way radios. If the lead hi-railer makes an emergency stop while on track, they are to inform the following vehicle immediately.
- Both [redacted] hi-railers will shunt the track to activate the gates for gated crossings. However there are can be numerous highway grade crossings without gates. At these crossings, the hi-railers will slow down to a safe speed (as determined by [redacted] personnel) or come to a full stop to clear the crossing before proceeding.

Data Management Plan

Participants will be assigned a unique number and each participant's data will be labeled with this number. The only permanent record of the individual's name will be on the consent forms. The consent forms do not contain the participant's number; therefore at no point can the individual's name be linked to the participant number. Consent forms are stored in a locked filing cabinet.

Collected data will be labeled with the participant number only. This data will be stored on a secure, password-protected QNA computer accessible only by the Principal Investigator. Participants will be informed that only the researchers have access to their data and that their identities will not be disclosed to outside parties in any form. No individual identifying information will be used in reports and/or publications.

Appendix C.
Consent Form



QinetiQ
North America

CONSENT TO PARTICIPATE IN RESEARCH STUDY

This form will give you important information about why this study is being conducted and what will happen during the study, including the risks and possible benefits. Please read it carefully. After you finish, the researcher will answer any questions that you may have.

1. **Overview of Study**
   Study Title: Hybrid Track Inspection Study
   Sponsor: Federal Railroad Administration, 1200 New Jersey Ave SE., Washington, DC 20590
   Principal Investigator (PI): Amanda DiFiore, QinetiQ North America, Inc., Technology Solutions Group (QNA), 358 Second Ave., Waltham, MA 02451
   Purpose of this Study: This study is being conducted determine the effectiveness of human track inspection as well as the effectiveness of automated track inspection.
   Information about Study Participants: You are being asked to take part in this study because you are member of the population of interest, i.e. track inspector or exception editor.

2. **Participation Requirements**
   Time commitment: You will participate in a 30 minute safety briefing, test procedure instruction and system calibration period, and then partake in an hour and a half long test session.

   Study Procedure:

   Data Collection at Track

   1. Upon arrival at track, you will be asked to read the consent form. Your participation is voluntary and can be ended at any time if you chose to no longer participate. All of the participant specific data collected will only be utilized for study purposes and will not be shared with FRA except anonymously or as statistical summary data. Your performance in this study will affect your job in no way. Please inform us of any privacy or safety concerns. Once all your questions are resolved and you are satisfied, please proceed and sign the consent form.

   2. After you have read and signed the consent form, the test will begin. Please inspect the track as you would normally to FRA class [TBD] standards according to 49 CFR 213.

   3. You will be a passenger in the hi-railer and can direct the driver to change speed, stop, or reverse when necessary to perform a thorough visual track inspection. However, the hi-railer cannot exceed track speed limits or safe operating speed limits for hi-rail vehicles. When you perceive that the hi-railer is within the zone of the POI, we ask that you press the button and verbally describe the POI. A POI can be of two different types: a single location on track (e.g broken tie), or a length of

track (e.g low ballast.) For the former, push the button when the hi-railer is at the location of the POI. For the latter, push the button at the beginning of the length of track where the POI exists. GPS location will be recorded with each button press to indicate the location of the POI.

4. In addition to the visual track inspection we will permit you to leave the hi-rail vehicle ONLY to conduct a walking inspection for switches. However, if the walking inspection of switches exceeds time allotted for the research study we will omit them. While the methodology of visual track inspection does not permit exiting the vehicle to verify your POI, you will have the opportunity later in the study to participate in an examination of all POIs identified using GPS location data. This ground verification will take place on Thursday and Friday.

Data Collection Off-Site with Exception Editors

1. Please read the consent form. Your participation is voluntary and can be ended at any time if you chose to no longer participate. All participant specific data collected will only utilized for study purposes and will not be shared with FRA or your employer except anonymously or as statistical summary data. Your performance in this study will affect your job in no way. Please inform the PI of any specific safety or privacy concerns. Once all questions are resolved you are satisfied, please sign the consent form.

2. The PI will give you instructions on how to operate the iPad application and how to record responses.

3. Once you are comfortable with the iPad application, you will be presented with a playback of the TGMS data feed/strip charts and asked to filter the exceptions using your normal decision making process to FRA class [TBD] standards. Please record responses using the iPad data collection application instead of your normal software

4. The test ends once you have reached the end of the data file.

3. **Compensation**

Track Inspectors: Test participants will be FRA track inspectors. Because the study will be conducted as part of your normal working hours, no special compensation for participation will be provided. Travel for the participants will be paid directly by FRA. A total of 12-15 track inspectors will participate. All FRA track inspectors are assumed to be experts with several years of experience.

Exception Editors: Test participants will be [redacted] employees. Because the study will be conducted as part of their normal working hours, no special compensation for participation will be provided. The filtering process will be conducted post data collection and off-site at an [redacted] location, and as such, travel costs are not anticipated. All [redacted] exception editors are assumed to be experts with several years of experience. A total of 12-15 editors will participate.

4. **Potential Risks & Research Related Injury Coverage**
None are expected. You are not performing any actions or are not involved in any processes that are not part of your normal working duties.

5. **Potential Benefits**
The data collected in this study will help the FRA develop better guidelines for track inspection. Better track inspection guidelines will result in the detection of defective track conditions earlier and quicker. Because most train derailments and railroad accidents are caused by defective track conditions, it is expected that the outcomes of this research will lead to a safety improvement for individuals working on or around railroad track.

6. **Voluntary Participation and Termination of Participation**
Your participation is voluntary and you may withdraw your consent to participate at any time without any negative effects. To terminate your participation inform the investigator of your desire to end the study. Any information recorded or collected pertaining to you will be immediately destroyed. As an employee of FRA or [redacted], you are under no obligation to participate in this study. Study participation is completely voluntary and withdrawal from the study can occur at any time for any reason. The decision to not participate in this research study, or a decision to withdraw from a study, will have no effect whatsoever on employment status at FRA or [redacted].

7. **Confidentiality and Privacy**
QNA will assign a unique participant number to each study participant. All collected data will be organized and referenced by participant number. No information will be provided to the railroads, the labor unions, FRA or [redacted] as to the names of the participants nor will individual data from a single participant be disclosed.

8. **Questions**
You may contact the QNA Investigator at any time with questions about the study or the conditions of your participation. The contact information is as follows:

Ms. Amanda DiFiore
Engineering Manager
QinetiQ North America, Inc
358 Second Avenue
Waltham, MA 02451
amanda.difiore@qinetiq-na.com
781-684-3978

If you have questions about your rights as a research subject, you may contact Richard Clunie, CEO, at Asentral, Inc. Institutional Review Board at 978-462-6415. Asentral, Inc. Institutional Review Board is an ethical review board that has reviewed and will oversee this study with your safety and welfare in mind.

9. Consent and Signature

I have read and understand the requirements of my participation in this study as described in this Informed Consent, as well as my right to refuse to participate or to terminate my participation at any time. I understand that the information collected will be kept confidential. My questions have been answered and I agree to voluntarily cooperate and participate in good faith.

Name(print): ____________________

Signature: ____________________ Date: __________

Witness: ____________________ Date: __________

Appendix D.
p(Hit) and *p*(FA) Data

This appendix provides the number of Hits, FAs, Misses, and CRs for each TI and ETG participant for each of the four observation conditions (MR, SC, TGMR, and TGSC). In addition, the tables provide the Hit rate and FA rate.

The RR1 test zone was approximately 11 miles (58,080 feet) long, and the RR2 field test was approximately 13 miles (68,640 feet) long. Dividing the total length of each test zone by 160 feet provides the number of 160-foot segment lengths in each test zone. There are 363 segments and 429 segments for the RR1 and RR2 test zones, respectively. The number of CRs was calculated by subtracting the sum of Hits, FAs, and Misses from the total number of 160-foot segment lengths. For example, for ETG1 in RR1 in Table 5, the total number of segments is 363 and subtracting the total number of Hits (15), FAs (1), and Misses (74) from 363 results in 273 CRs.

Table 5. Data from RR1 field test for MR condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	15	1	74	273	0.169	0.004
ETG 2	16	0.5[9]	73	274	0.180	0.002
ETG 3	15	0.5[10]	74	274	0.169	0.002
ETG 4	17	4	72	271	0.182	0.015
TI 1	21	6	67	269	0.239	0.022
TI 2	19	3	69	272	0.216	0.011
TI 3	43	9	32	267	0.573	0.033
TI 4	31	3	56	273	0.356	0.011

Table 6. Data from RR2 field test for MR condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	17	4	114	294	0.130	0.013
ETG 2	17	3	114	295	0.130	0.010
ETG 3	14	2	117	296	0.107	0.007
ETG 4	9	1	123	297	0.068	0.003
TI 1	77	13	51	286	0.602	0.043
TI 2	44	19	85	281	0.341	0.063
TI 3	26	3	105	295	0.198	0.010
TI 4	83	8	48	290	0.634	0.027

---

[9] ETG 2 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

[10] ETG 3 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

Table 7. Data from RR1 field test for SC condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	15	1	31	316	0.326	0.003
ETG 2	16	0.5[11]	30	317	0.348	0.002
ETG 3	15	0.5[12]	31	317	0.326	0.002
ETG 4	16	4	30	313	0.348	0.013
TI 1	12	15	33	303	0.267	0.047
TI 2	9	13	36	305	0.200	0.041
TI 3	23	36	17	282	0.575	0.113
TI 4	18	16	26	303	0.409	0.050

Table 8. Data from RR2 field test for SC condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	13	8	17	391	0.433	0.020
ETG 2	14	6	16	393	0.467	0.015
ETG 3	12	4	18	395	0.400	0.010
ETG 4	8	2	22	397	0.267	0.005
TI 1	20	69	9	328	0.690	0.174
TI 2	14	49	17	349	0.452	0.123
TI 3	9	20	21	379	0.300	0.050
TI 4	20	71	10	328	0.667	0.178

---

[11] ETG 2 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

[12] ETG 3 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

Table 9. Data from RR1 field test for TGMR condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	15	1	35	312	0.300	0.003
ETG 2	16	0.5[13]	34	313	0.320	0.002
ETG 3	15	0.5[14]	35	313	0.300	0.002
ETG 4	17	3	33	310	0.340	0.010
TI 1	16	11	34	302	0.320	0.035
TI 2	7	15	42	299	0.143	0.048
TI 3	26	33	18	281	0.591	0.105
TI 4	21	13	27	302	0.438	0.041

Table 10. Data from RR2 field test for TGMR condition

	Hits	FAs	Misses	CRs	*p*(Hit)	*p*(FA)
ETG 1	17	4	21	385	0.447	0.010
ETG 2	18	2	21	387	0.462	0.005
ETG 3	15	1	21	388	0.417	0.003
ETG 4	9	1	22	388	0.290	0.003
TI 1	26	62	7	324	0.788	0.161
TI 2	15	48	26	340	0.366	0.124
TI 3	14	15	15	374	0.483	0.039
TI 4	23	68	17	321	0.575	0.175

---

[13] ETG 2 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

[14] ETG 3 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

Table 11. Data from RR1 field test for TGSC condition

	Hits	**FAs**	**Misses**	**CRs**	***p*(Hit)**	***p*(FA)**
ETG 1	15	1	23	324	0.395	0.003
ETG 2	16	0.5[15]	22	325	0.421	0.002
ETG 3	15	0.5[16]	23	325	0.395	0.002
ETG 4	17	3	21	322	0.447	0.009
TI 1	11	16	27	309	0.289	0.049
TI 2	7	15	30	311	0.189	0.046
TI 3	19	41	14	285	0.576	0.126
TI 4	15	19	21	308	0.417	0.058

Table 12. Data from RR2 field test for TGSC condition

	Hits	**FAs**	**Misses**	**CRs**	***p*(Hit)**	***p*(FA)**
ETG 1	13	8	7	401	0.650	0.020
ETG 2	14	6	6	403	0.700	0.015
ETG 3	12	4	8	405	0.600	0.010
ETG 4	8	2	12	407	0.400	0.005
TI 1	12	76	7	330	0.632	0.187
TI 2	10	53	11	355	0.476	0.130
TI 3	7	22	14	386	0.333	0.054
TI 4	12	79	8	330	0.600	0.193

---

[15] ETG 2 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

[16] ETG 3 had no FAs. Having a value of 0 or 1 for *p*(Hit) or *p*(FA) prevents the use of some standard SDT equations. Therefore, a value of 0.5 was used for the number of FAs instead of a value of 0. This prevents *p*(FA) from having a value of 0.

Abbreviations and Acronyms

CFR	Code of Federal Regulations
CR	Correct Rejection
ETG	Edited Track Geometry
FA	False Alarm
FRA	Federal Railroad Administration
GRMS	Gage Restraint Measurement System
MR	Maintenance Required
POI	Point of Interest
SC	Safety Critical
SDT	Signal Detection Theory
TGMR	Track Geometry Maintenance Required
TGMS	Track Geometry Measurement System
TGSC	Track Geometry Safety Critical
TI	Track Inspector
TSS	Track Safety Standards
RR1	Railroad segment for first field test
RR2	Railroad segment for second field test