Assessment of Rail Flaw Inspection Data

Brandon D. Jeffrey
M. L. Peterson

Department of Civil Engineering
Colorado State University
Fort Collins, Colorado

August 1999

ACKNOWLEDGMENTS

It is an honor to thank the following individuals for their support, help, and guidance. A first thanks is given to Alex Trindade, Ph.D. candidate at Colorado State University in statistics, for his expertise in statistical analysis. Also, in the statistics department at Colorado State University, a Master's candidate Jill Smith, proved to be a tremendous help. Special thanks is given to Dr. Michael Peterson, professor of mechanical engineering at Colorado State University, for his guidance throughout this project. Finally, for the real world experience, Greg Garcia of the Transportation Technology Center Inc., for his time and efforts in providing a unique educational experience that will carry forward to a career.

DISCLAIMER

ABSTRACT

This project is an analysis of the rail flaw data for the Transportation Technology Center Inc. (TTCI). Six major railroads provided TTCI with defective rail, that had been removed from service. The rail, which contained a variety of defects, was placed into a gauntlet track at the Transportation Technology Center (TTC). The gauntlet track is known as the Railflaw Detection Test Facility (RDTF) and is a railroad industry tool, in cooperation with the Federal Railroad Administration (FRA), to evaluate rail flaw detection technology. TTCI inspected the rail, catalogued the defects, and installed the rail in the RDTF. There are 49 defects for each inspection vehicle evaluation. To date, there have been six evaluations performed at the RDTF. This paper is the assessment of those six evaluations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

1. INTRODUCTION

The detection of defects in steel rail is perhaps the first widespread application of non-destructive testing. Failure to detect defects is significant for the development of non-destructive testing and, more importantly, the safety of the railroads. However, in the decades that have passed since the initial development of the rail defect detecting technology, the basic inspection processes have not changed. Significant developments have been made in signal processing and automated rail flaw evaluation. This project is a small part of an assessment of the rail inspection industry and the many factors that contribute to the ability to detect defects in steel rail.

The Transportation Technology Center Inc. (TTCI), a wholly owned subsidiary of the Association of American Railroads (AAR), and the Federal Railroad Administration (FRA) have cooperated to explore current technologies for rail inspection with an emphasis on reducing risk, while maintaining cost control on the maintenance of the railroads. TTCI is partially funded by the six major railroads whose rail lines represent a significant portion of the track mileage and the total tonnage, which is transported by rail. The primary technical efforts will focus on detection of typical defects that occur in rails. As an experimental base, examples of true defects actually have been removed from revenue service lines and placed in track section in a test loop. The test loop is located on the 52-square-mile test center at the Transportation Technology Center (TTC), east of Pueblo, Colo. The defects are hosted in conventional curved and tangent track segments. In addition, fabricated defects were placed in the track sections for determining the limitations to the testing technologies.

Support for this research paper has been provided in part by the U.S. Department of Transportation (USDOT) via the Mountain Plains Consortium, which is sponsored through the University Transportation Centers Program. The efforts of Colorado State University (CSU) have concentrated on providing an assessment of evaluations of the rail inspection vehicles. Through the spring of 1998, TTCI has tested six rail inspection vehicles to provide detection results for the 49 catalogued defects on the gauntlet course. The results presented in this paper consist only of results from the first six trials, the statistics are expected to change as additional cars are tested. This paper provides an overview of the ultrasonic inspection technology, the gauntlet course, plus its defects, and the statistical results from six evaluations of the rail inspection vehicles.

2. CURRENT AND PAST TECHNOLOGIES

History

Non-destructive testing had its beginning in rail application in 1923 when Dr. Elmer Sperry noticed an increase in the number of disastrous train derailments. Sperry developed the first rail inspection car (Car SRS #101) to detect transverse fissures in railroad rails. By November 1928, Sperry had an inspection car testing rail on the Wabash Railway in Montpelier, Ohio, and Clarke Junction, Ind., for its first commercial use. Car SRS 102 detected a large transverse fissure in the head of a rail. Much to the annoyance of the railroad, the rail was taken out of service and the following day was tested by Sperry's chief research engineer. The rail was tested and broken where the transverse fissure was found. After the convincing test of the SRS 102, Sperry expanded his services and increased his fleet to 10 cars by the end of 1930 (Clarke, 1997).

The first inspection car employed an induction method. A heavy current was induced through the rail to be tested. Search coils were moved through the resulting magnetic field to find perturbations in this field caused by defects. By 1959, Sperry developed the first ultrasonic test car for the New York City Transit Authority (NYCTA). Today, Sperry has a fleet of more than 40 test cars, both, magnetic induction cars and ultrasonic detection cars that employ both methods.

By the 1950's track conditions significantly improved when continuous welds of each rail replaced joint bars. This resulted in a shift of defect types found in the rails, away from joint defects. Over the next 10 to 20 years the average age of the rail naturally increased as did the average load of the rail cars. Then, in the 1980's, with increased use of inspection cars throughout North America the general trend in detection was lower defect rates (Davis, 1997).

Today, interest in repair and maintenance, and the characteristics of the railroad industry have changed. There is great pressure on operations to gain efficiencies for greater financial returns with increased traffic. This pressure has resulted in heavier axle loads and higher train speeds. Due to higher train speeds, shorter work windows exist for the conducting of ultrasonic inspection. Fifteen to 20 years ago, a typical rail life was 800 million gross tons. Today, 1.5 billion gross tons is not unusual for a 136-lb. rail (136-lb. is the force at which the rail undergoes yield deformation and is the standard rail rating). Federal regulations require immediate remediation of detected defects, regardless of the type size or the amount of traffic the rail has seen. It has been the industry standard to replace rail when six to eight defects are detected per mile per year. In the past, it was four defects per mile per year (Franke, 1997).

A fleet of 30 detector cars can cover about 110,000 miles per year. On the average, every mile of track is inspected by a rail detector car at least three times a year. Table 2.1 is a list of the type of derailments attributed to rail defects from 1992 to 1995. The breakdown of cost associated with the derailments also is listed.

Table 2.1: Rail Derailments, 1992–1995

Cause	1992	1992 Costs	1993	1993 Costs	1994	1994 Costs	1995	1995 Costs	Total	Total Costs
Bolt Hole Crack	3	$329,214	7	$1,369,414	8	$1,847,900	9	$1,922,651	27	$5,469,179
Broken Weld	3	$1,063,252	3	$510,300	4	$772,485	7	$1,416,539	17	$3,762,576
Detail Fracture	15	$4,562,548	12	$4,950,399	17	$3,789,256	14	$1,554,079	58	$14,856,282
Head/Web	25	$1,147,756	31	$2,276,856	25	$1,308,153	30	$2,885,749	111	$7,618,514
HSH	7	$236,463	8	$356,833	5	$335,558	6	$412,825	26	$1,341,679
Broken Bars	9	$803,585	6	$514,240	9	$1,082,553	6	$1,207,444	30	$3,607,822
TD	37	$3,057,747	46	$5,779,296	32	$5,277,520	44	$7,251,921	159	$21,366,484
VSH	24	$1,222,085	35	$3,984,662	40	$5,914,481	27	$1,815,491	126	$12,936,719
Totals	**123**	**$12,422,650**	**148**	**$19,742,000**	**140**	**$20,327,906**	**143**	**$18,466,699**	**554**	**$70,959,255**

(Data is from Federal Railroad Administration "Accident/Incident Bulletin", 1992 –1995)
Notes: Broken base and broken bolt derailments are not included, since they cannot be detected by ultra-sonic detector cars.
HSH – Horizontal Split Head TD – Transverse Defect VSH – Vertical Split Head

Of the defects listed, detail fractures and transverse defects represent about 40 percent of the causes of train derailments during 1992-95 and 51 percent of the total costs. Today's world of limited resources requires strict management of railroad assets. One key asset for the railroads are the rails. The rail industry invests millions per year in rail detection alone, which is key to preventing failures. From the

Burlington Northern and Santa Fe Railway (BNSF) Chief Engineer M. W. Franke provided a most interesting philosophy, "…every dollar spent on rail detecting is worthwhile even if it were to prevent only one major derailment per year, the rail detection program has more than paid for itself..." This approach ensures safety of the rails and maintains high efficiency for provision of services. The greatest challenge however, is to maintain the required rail inspection schedule given tight scheduling requirements, particularly for main line traffic.

Ultrasonic Technology

Ultrasonic inspection technology is the predominant rail inspection technology used in North America. Most of the major contractors rely exclusively on ultrasonic rail testing cars (see Figures 2.1 and 2.2). Some companies do employ larger rail bound units that have ultrasonic and magnetic induction technologies (see Figure 2.1). Currently, the ultrasonic rail testing cars are capable of speeds up to 30 mph in a non-stop testing mode. However, in practice the vehicle stops frequently to hand verify indications from the rail testing. With the stopping and verification rail inspection vehicles typically average 6 to 8 mph in practice.

Figure 2.1: Road / Rail Ultrasonic Car

(Picture of Sperry Rail Service Car SRS #821 taken by Al Bowen.)

Figure 2.2: Railbound Induction / Ultrasonic Car

(Picture of Sperry Rail Service Car SRS #143 taken by William D. Miller)

The ultrasonic system traditionally is a pulse-echo method where standard ultrasonic piezo-electric transducers are mounted into wheels filled with fluid (Garcia, 1998). The wheels rolling along the track send signals through the rail to be received by the transducer. To transmit the ultrasonic signal from the wheel to rail, a coupling media is needed. A coupling fluid is applied between the wheel and rail contact area and usually is oil- or water-based.

A transducer orientation in the wheel is fixed to offer different inspection capabilities. Different transducer angles are used to find different types of defects. Inside each wheel three to sex transducers are mounted, usually with two wheels per rail. A total of as many as 12 transducers per rail are used. Transducers commonly have four different orientations: 0^o, 45^o, and 70^o from the vertical, and a "side looker" transducer may be located in each wheel (see Figure 2.3). The 0^o transducer is mounted, perpendicular to the track, and picks up horizontal defects. The 45^o and 70^o transducers are mounted in a forward and reverse position in the wheel, which allows for detection of defects in various orientations in the rail. The "side looker" transducers are sensitive to vertical and shear type defects. The angles mentioned above are not the directions of the transducer in the wheel, but the angle of the transmitted longitudinal wave in the steel rail, known as the refracted angle. The direction of the transducer in the wheel is known as the angle of incidence and can be calculated from Snells Law of Refraction:

The Law of Refraction

$$\text{Sin}(\beta) = (V_1 / V_2)\ \text{Sin}(\alpha)$$

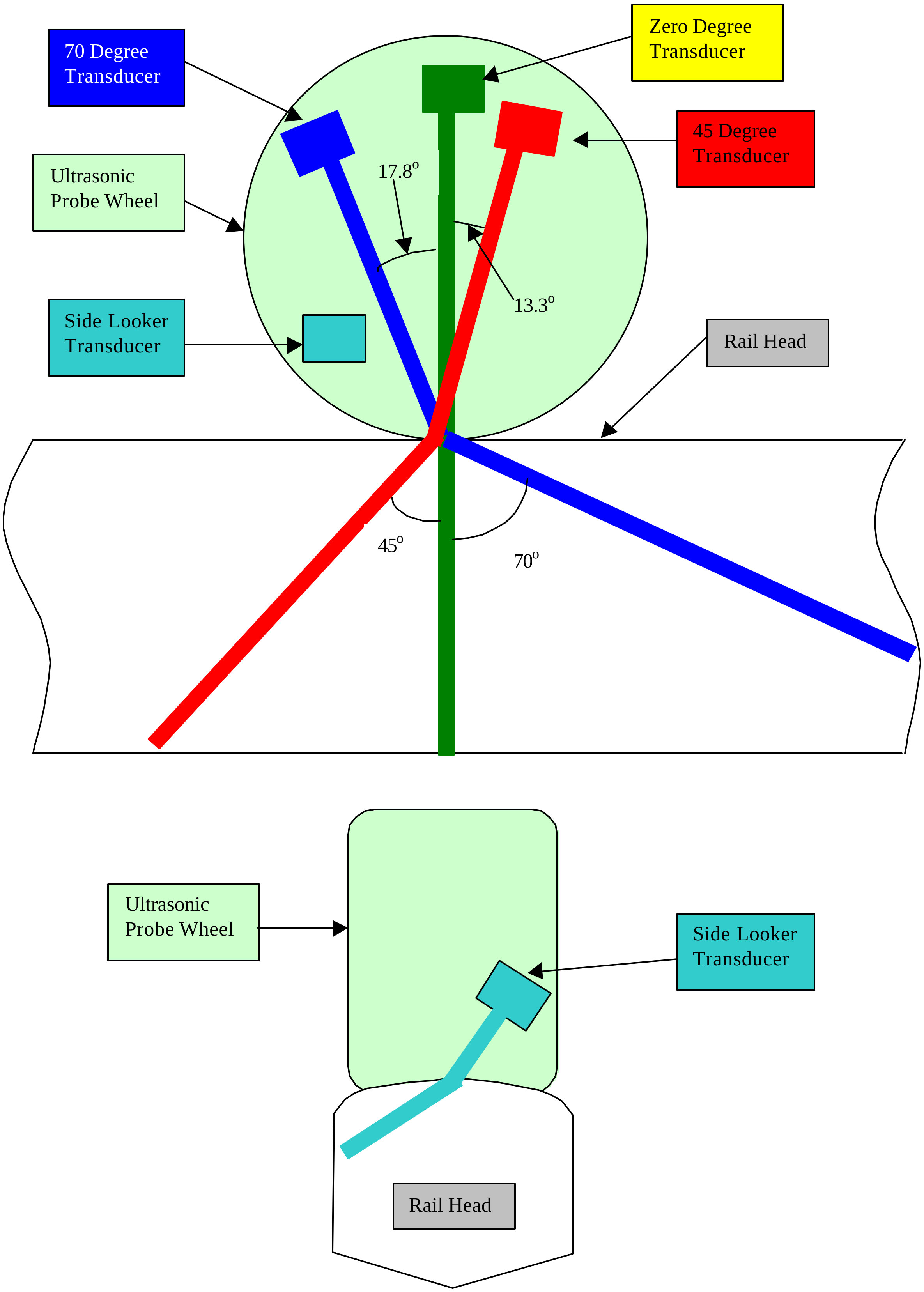
where: α = angle of refraction/emergence (angle of transmitted longitudinal wave)
β = angle of incidence (angle of transducer in wheel)
V_1 = velocity of sound wave in wheel fluid (Glycerin) = 1.92 km/s
V_2 = velocity of the longitudinal sound wave in rail (Iron) = 5.9 km/s

A few simple calculations determine the angles at which the transducers are arranged in the wheel. Table 2.2 summarizes the common angles of incidence and the angles of refraction.

Table 2.2: Ultrasonic Wave Angles

a = Angle of refraction/emergence	**b = Angle of incidence**
0 degree transducer (angle in rail)	0.0 degrees (angle in wheel from vertical)
45 degree transducer (angle in rail)	13.3 degrees (angle in wheel from vertical)
70 degree transducer (angle in rail)	17.8 degrees (angle in wheel from vertical)

Figure 2.3: Ultrasonic Probe Wheel Transducer Arrangements



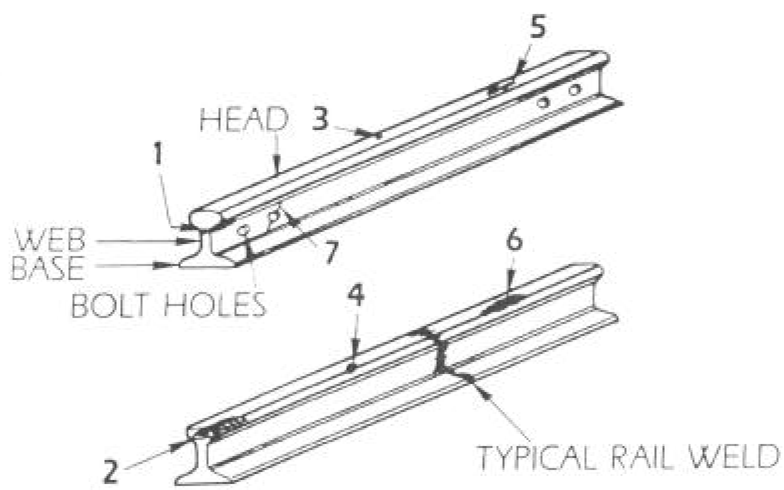
Research Overview

The rail inspection system for finding defects currently has some shortcomings. First, as a traditional ultrasonic system, the operator is required to set the threshold value for calibration. Calibration of the ultrasonic equipment is used to count pixels to define a defect size. The operator must watch the display screen signal to recognize defects. If a particular signal appears to represent a defect, the vehicle will stop and the indication will be hand verified using a flaw detector. This requires the inspection vehicle to stop and back up to verify the defect. Ultimately, the operator must make all the evaluations and all detected defects must be verified. This process is time consuming and can create scheduling problems for the railroad.

In the inspection system used in North America, limits on test productivity exist. The inspection test vehicle only is allowed to operate until the number of defects found can be handled by rail replacement forces crew gangs. A replacement crew gang can repair 10 to 20 miles of track per day, which limits test productivity.

Most notably, current data indicates that 70 to 80 percent of rail detector car indications prove false during hand test verifications. Significant time is consumed hand verifying false indications. Another concern is that current detector cars systems fail to detect detail fractures or other defects masked by shelled rail, detail fractures under spall, or even dirty rail.

Figure 2.4: Typical Rail Defects



(From the U.S. Dept. of Transportation "Accident/Incident Bulletin", 1984)

Notes: 1 – Head and Web Separation
2 – Horizontal Split Head
3 – Detail Fracture
4 – Transverse Fracture
5 – Engine Burns
6 – Shelling
7 – Bolt Hole Cracks

In 1996 there were six main line derailments from detail fractures that were not detected during inspection. These defects were missed because the defect was located under a shelled rail surface (Zarembski, 1997). Thus, the frequency of detecting false indications has to be reduced while improving the ability to detect certain types of defects.

Needs that currently are being addressed by the technical community are training, quality assurance, maintenance, and new technology. With current testing speeds of 6-8 mph the need for faster testing is a priority. Later this year it is expected that a new system of high-speed non-stop operations testing will be adopted. This is to be done by linking a high-speed test car via modem with a chase car that is used to perform hand verification. This approach will dramatically reduce track congestion by speeding up the testing process. Chase cars will be able to receive digitized analog data transferred from the defect inspection vehicle. The chase car will stop at indications and hand verifications of the detected defects will be performed. This new chase car system is projected to run at a 50 mph test speed between stops and would double the miles tested per hour (Zarembski, 1997). Also being proposed is a new computer system with digital processing. It is designed to reduce the interpretation burden for the operator and will be able to identify defects to provide additional information to the operator. Most importantly, recent development work is focusing on improving the testing for vertical split heads (VSH) and transverse defects (TD) under spall or shelling (these defects are described in further detail in Chapter 3, Section C).

3. GAUNTLET TEST COURSE

Gauntlet Description

The 52-square mile Transportation Technology Center (TTC) offers a location for evaluating and testing railroad-related equipment and track safety improvements. Innovations are taken from the drawing board to the test track. The joint AAR and FRA project, which produced the gauntlet test course, is a research program exploring methods to detect known rail defects. TTC has 48 miles of railroad track on-site devoted to testing consisting of three test loop sections. Connected to the largest test section the Railroad Test Track (RTT), is a balloon loop, which has a seven-degree 30-minute curve with 4.5 inches superelevation and a five-degree reverse curve with 3.5 inches of superelevation. The balloon loop is the site of the Railflaw Detection Test Facility (RDTF).

Figure 3.1: Railflaw Detection Test Facility (RDTF)

(From TTCI of the RDTF ‘gauntlet course’, 1998)

Rail has been supplied by the six major railroads to provide TTCI with 56 flawed rail sections with various types of internal and surface anomalies. These 56 samples were used to create a facility for evaluating effectiveness and efficiency of current rail inspection vehicles such as those shown in Figures 2.1 and 2.2. Initial evaluations performed on the RDTF were conducted to benchmark the performance of current railflaw technology. This will allow for performance comparisons of the ultrasonic sensor technology. TTCI inspected the rail sections prior to installation, catalogued the defects, and joined the flawed rail into curved and tangent track segments, which are connected to the balloon loop for the RDTF. A consensus between TTCI, the railroads, and the inspection suppliers was made on which defects would be used in the railflaw benchmarking evaluations. Altogether 49 defects, mainly consisting of detail fractures (see Table 3.1), were selected. Thus, the rail detection test facility (RDTF) will be used to evaluate the current ultrasonic detection systems and ultimately to further the development of improved railflaw technology.

Table 3.1: Gauntlet Course Defect List

WEST/HIGH RAIL

	DEFECT	SIZE
1	DF	3%
2	DF	6%
3	DF	24%
4	DF	68%
5	DF	4%
6	DF	5%
7	DF	45%
8	DF	33%
9	DF	23%
10	DF	17%
11	DF	14%
12	DF	25%
13	DF	19%
14	DF	30%
15	DF	5%
16	DF	40%
17	DF	48%
18	DF	73%
19	DF	3%
20	DF	5%
21	DF	42%
22	DF	8%
23	DF	20%
24	DF	6%
25	DF	18%
26	DF	6%
27	DW	5%
28	DW	6%
29	DW	11%
30	HSH	3"X2"
31	BHC	0.25
32	BHC	0.38

EAST/LOW RAIL

	DEFECT	SIZE
33	DF	22%
34	DF	10%
35	DF	14%
36	DF	14%
37	DF	4%
38	DF	15%
39	DF	5%
40	DF	8%
41	DF	11%
42	DF	25%
43	DF	10%
44	DF	3%
45	DF	21%
46	DF	8%
47	DW	13%
48	VSH	120"
49	HSH	2"X1"

(Defect listing is from TTCI for Vendor Inspection Vehicle Evaluation, 1998)

Notes: Defect type: DF – Detail Fracture
BHC – Bolt Hole Crack
DW – Defective Weld
VSP – Vertical Split Head
HSP – Horizontal Split Head
Size is the percent of cross sectional head area for a TD or DF that has fractured or is in question.

Rail Defect Documentation

The 56 flawed rails were subjected to a series of non-destructive tests to document the defects in the rail before it was joined in track. The rail was supplied by railroads with the rail containing defects. TTCI selected samples and inspected the flawed rail before installation into RDTF.

- **Visual Inspection**

Figure 3.2: Industry Donated Rail

(Rail samples used in RDTF located at TTC, TTCI 1998)

- **Ultrasonic Hand Inspection**

Figure 3.3: Ultrasonic Testing of Rail Samples

(Rail sample used in RDTF located at TTC, TTCI 1998)

- **Radiographic Inspection**

Figure 3.4: Radiography of Donated Rail Samples

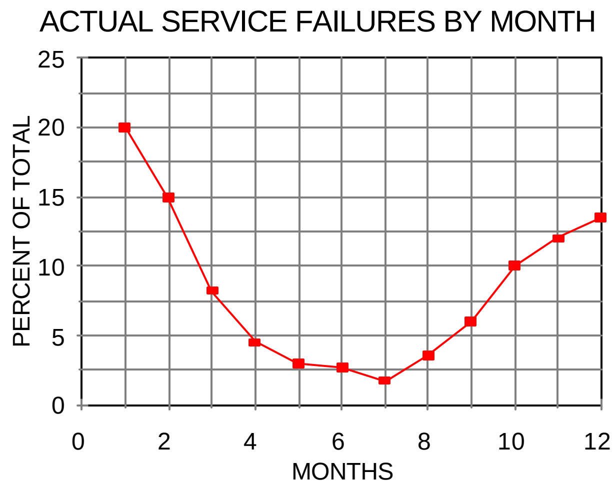
(Rail samples used in RDTF located at TTC, TTCI 1998)

The rail received at TTC was inspected visually, ultrasonically and, in some cases, radiographically. The inspections were performed to document the external and internal condition of the rail. The visual inspections included identification of surface conditions such as shelling, spalling, head checks, and rail wear. The ultrasonic inspection consisted of ultrasonic hand-mapping of the head, web and base using pulse echo A-scans with 0, 45 and 70 degree transducers. Radiography of the rail was performed only on rail containing transverse defects and was performed to document the orientation of the defect when referencing the top of the rail on the gage face side (Garcia, 1998). TTCI used these methods of inspection before installing the flawed rail into the RDTF.

Types of Rail Defects

The rail defects on the following pages show the most common defects of concern to the industry. It has been shown that defects that cause train derailments most frequently occur in the winter months (see Graph 3.1). In the winter months the rail track goes into a tensile behavior and produces 64 percent of the detected defects in the six coldest months of the year and 75 percent of service failures. Winter also is when the rail inspection equipment reliability is the lowest (Davis, 1997).

Graph 3.1: Service Failures From 1986-1988



(A Study of the Seasonality of Rail Defect Occurrence, TTCI 1997)

Types of rail defects the railroad industry focuses on detecting to prevent service failures are shown in Figure 3.1. The following figures are the typical defect in the RDTF.

Figure 3.5: Detail Fracture (DF)

(From TTCI FAST Program, 1993)

Figure 3.5 is a Detail Fracture (DF) and is the most common defect in the RDTF. The DF has an origination point and grows radially from the origination point. These types of defects are caused from excessive stress concentrations.

Figure 3.6: Transverse Defect (TD)

(From the U.S. Dept. of Transportation “Accident/Incident Bulletin,” 1984)

Figure 3.6 is a Transverse Defect (TD) and is the most critical type of defect, causing 29 percent of train derailments. Transverse defects have an origination point in the center of the fissure and they grow circular to the origination point. A TD is caused from fatigue of the rail.

Figure 3.7: Vertical Split Head (VSH)

(From the U.S. Dept. of Transportation “Accident/Incident Bulletin,” 1984)

The vertical split head causes the second most train derailments at 23 percent of total derailments (Figure 3.7). Vertical split heads usually originate from manufacturing anomalies. For the defect that has caused 20 percent of train derailments from Table 2.1, is the head and web separation shown in Figure 3.8. Head and Web separations often are caused from excessive stress concentration.

Figure 3.8: Horizontal Split Head (HSP)

(From the Accident/Incident Bulletin No. 153 of the U.S. Dept. of Transportation, 1984)

The rest of the defects – detail fracture, horizontal split head, bolt hole cracks, and shelling – represent the other 28 percent of the defects causing train derailments during 1992 through 1995 of Table 2.1. A current concern in the industry is that current inspection methods are not detecting defects under shelling or defects masked by spalled rail.

Figure 3.9: Detail Fracture Under Shelling

(From TTCI FAST Program, 1993)

Figure 3.9 is a detail fracture under shelling. The detection of these defects still needs to be investigated.

Railflaw Evaluations on the RDTF

Once the industry-donated rail was non-destructively inspected and installed into the RDTF, TTCI performed benchmarking evaluations of inspection technologies. These evaluations were performed to provide the benchmarking of current railflaw ultrasonic technology. There have been six evaluations performed to date. Results of those evaluations are listed in Table 3.2.

Table 3.2: Railflaw Evaluation Results

	RAIL	EFFECT	SIZE	E1	E2	E3	E4	E5	E6
1	H/W	DF	3%	Y	N	Y	N	N	N
2	L/E	DF	22%	Y	Y	Y	Y	Y	Y
3	L/E	DF	10%	Y	Y	Y	Y	Y	Y
4	L/E	DF	14%	Y	N	N	Y	N	N
5	L/E	DF	14%	Y	N	N	Y	Y	N
6	L/E	DF	4%	Y	N	N	Y	N	N
7	L/E	DF	15%	Y	N	N	Y	N	Y
8	L/E	DF	5%	Y	N	Y	Y	Y	Y
9	H/W	DF	6%	Y	Y	Y	Y	Y	Y
10	H/W	DF	24%	Y	Y	Y	Y	Y	Y
11	H/W	DF	68%	Y	Y	Y	Y	Y	Y
12	H/W	DF	4%	N	Y	N	Y	Y	N
13	H/W	DF	5%	N	Y	N	Y	Y	Y
14	H/W	DF	45%	Y	Y	Y	Y	Y	Y
15	H/W	DF	33%	Y	Y	Y	Y	Y	Y
16	H/W	DF	23%	N	Y	Y	N	Y	Y
17	H/W	DF	17%	N	Y	Y	Y	Y	Y
18	H/W	DF	14%	N	Y	Y	Y	Y	Y
19	H/W	DF	25%	N	Y	Y	Y	Y	Y
20	L/E	DF	8%	Y	Y	N	Y	Y	N
21	L/E	DF	11%	Y	Y	Y	Y	Y	Y
22	L/E	DF	25%	Y	Y	Y	Y	Y	Y
23	H/W	DF	19%	Y	Y	Y	Y	Y	Y
24	H/W	DF	30%	Y	Y	Y	Y	Y	Y
25	H/W	DF	5%	Y	N	N	N	N	N
26	H/W	DF	40%	N	Y	Y	Y	Y	Y
27	L/E	DF	10%	Y	Y	Y	Y	Y	Y
28	H/W	DF	48%	Y	Y	Y	Y	Y	Y
29	H/W	DF	73%	Y	Y	Y	Y	Y	Y
30	H/W	DF	3%	N	Y	N	N	N	Y
31	H/W	DF	5%	N	Y	N	N	Y	Y
32	H/W	DF	42%	N	Y	Y	N	Y	Y
33	H/W	DF	8%	N	N	Y	N	N	Y
34	L/E	DF	3%	Y	Y	Y	Y	Y	Y
35	H/W	DF	20%	Y	Y	Y	N	N	Y
36	H/W	DF	6%	N	Y	Y	N	Y	Y
37	H/W	DF	18%	N	Y	Y	N	Y	Y
38	L/E	DF	21%	Y	N	Y	Y	Y	Y
39	L/E	DF	8%	N	Y	Y	Y	Y	Y
40	H/W	DF	6%	N	N	Y	N	Y	Y
41	H/W	DW	6%	N	Y	N	Y	N	N
42	H/W	DW	11%	Y	N	N	N	Y	Y
43	L/E	DW	13%	Y	Y	Y	N	Y	Y

	RAIL	EFFECT	SIZE	E1	E2	E3	E4	E5	E6
44	H/W	DW	5%	N	Y	N	Y	Y	N
45	L/E	HSH	2"X1"	Y	Y	Y	Y	Y	N
46	H/W	HSH	3"X2"	Y	Y	Y	Y	Y	Y
47	L/E	VSH	120"	Y	Y	Y	Y	Y	Y
48	H/W	BHC	0.25	Y	N	N	Y	N	N
49	H/W	BHC	0.38	Y	N	N	Y	N	N

(Table is from TTCI railflaw evaluation results as of April 24, 1998)

Notes: Rail: L/E is Low / East and H/W is High / West
E1 – E6: are Evaluations 1 – 6.
Y – Defect detected
N – Defect not detected

Gauntlet Defect Study

A study of the evaluation results from Table 3.2 (defects frequently missed during benchmarking efforts) was performed. For this statistical evaluation a defect missed three or more times has been classified as a frequently missed defect. Table 3.3 was prepared to determine any associations between the frequently missed defects. In particular the defects size, orientation and/or rail surface condition.

Table 3.3: Frequently Missed Defects during Benchmarking Evaluations at TTC

#	Rail	Defect Size	Angle from RT	Rail Head Surface Condition
1	E/L	15% DF	Unknown	DF under spall
2	E/L	4% DF	Unknown	DF under shell
3	E/L	15% DF	Unknown	No surface anomalies
4	W/H	4% DF	Unknown	No surface anomalies
5	W/H	5% DF	Unknown	No surface anomalies
6	W/H	5% DW	Unknown	Flaking to slivering on head
7	W/H	23% DF	-3 degrees	No surface anomalies
8	W/H	5% DF	+2 degrees	No surface anomalies
9	W/H	3% DF	+4 degrees	No surface anomalies
10	W/H	5% DF	-10 degrees	No surface anomalies
11	W/H	42% DF	-5 degrees	No surface anomalies
12	W/H	8% DF	0 degrees	No surface anomalies
13	W/H	6% DW	Unknown	No surface anomalies (weld)
14	W/H	11% DW	Unknown	No surface anomalies (weld)
15	W/H	6% DF	Unknown	No surface anomalies
16	W/H	18% DF	+8 degrees	No surface anomalies
17	W/H	BHC	Unknown	No surface anomalies
18	W/H	BHC	Unknown	No surface anomalies
19	W/H	6% DF	+7 degrees	DF under shell

(TTCI gauntlet course defect comparison, 1998)

Notes: RT stands for radiographic inspection.
E/L is East / Low rail and W/H is the West / High rail.
Unknown means radiography unsuccessful or not performed to correlate orientation of defect.

An objective of the study was to find common orientations of missed defects. However, only eight defects listed in Table 3.3 had been subjected to radiograph tests to determine the defect orientation. Another objective was to find defects masked by surface anomalies. Only four defects were determined to be under surface anomalies. Because the sample sizes are small for these comparisons – defect orientations and railhead surface condition – no statistical evaluations were performed.

4. ANALYSIS OF RESULTS

Results

Track evaluations from the first six tests in Table 3.2 provide data used to evaluate the ultrasonic inspection vehicles. The data is comprised only of the success and failure of detecting a defect of certain size/type. With these results statistical analysis to determine significant relationships in the data was performed. The success and failure results of the 49 defects of the RDTF provide a basic understanding of the detection ratio, for each evaluation. Out of the 49 defects from Table 3.1, 44 were transverse in nature and are measured and sized in a similar fashion. The defects, which are similar, are detail fractures, transverse defects, and defective welds. The remaining defect classifications do not use the cross-sectional defect size. Because of this similarity, further statistical analysis was performed only on the 44 defects for which common sizing is applicable. The other defects were not included in the statistical evaluations. Two different statistical tests were performed to study a possible association between the size of the defect relative to the evaluation performed. This analysis was performed using the "chi-square test" and a relatively new procedure, the "logistic regression test." Further study of the results from the two statistical tests showed strong evidence of an association of failing to detect a defect of a given size only if all evaluations were combined. Therefore, using a larger sample size provided the only reliable statistical measure of determining if an association existed between defect size and success of detecting a defect.

Detection Ratios

The detection ratios for each of the evaluations are shown in Table 4.1. Significant variation in the results were expected due to conditions of operation. This is a source of error in the overall calculation of the probability of detection, but it is representative of actual field-testing. No one company inspects the entire rail system. The combined efforts of all of the vendors are used to detect flaws in the rail system. Therefore, the overall evaluation of the benchmarking tests provides the largest sample size and produces the most important results. The combined efforts of all vendors ensure the safety of the rail system. The detection ratios for all defects included in the railflaw evaluations are shown in Table 4.1. The additional detailed statistical analysis was performed on the 44 similar defects and is shown in the following sections.

Table 4.1: Detection Ratios for RDTF Evaluations

	POSSIBLE	ACTUAL	PERCENTAGE
EVALUATION 1	49	32	65.3%
EVALUATION 2	49	36	73.5%
EVALUATION 3	49	34	69.4%
EVALUATION 4	49	36	73.5%
EVALUATION 5	49	38	77.6%
EVALUATION 6	49	37	75.5%
OVERALL	**294**	**213**	**72.4%**

Chi-Square Test Statistic

The chi-square test is used to determine if size versus the success of detection is independent using a hypothesis method of association (Trindade, 1998). The chi-square test statistic was performed using both Minitab and SAS statistical analysis programs. To perform these two tests the data was split into different size categories.

The categories are as follows:

0 – 10% (small defect category)
11 – 20% (mid-small defect category)
21 – 30% (mid-large defect category)
31 – up (large defect category)

Hypothesis test:

H_o = Flaw size is independent of success rate
H_a = These classifications are dependent

If the size classifications are independent, H_o is true. This gives a p-value, which indicates the strength of the dependency between our two classifications (H_o and H_a). A smaller p-value increases the support of the alternative hypothesis, which indicates a stronger dependency between the two variables.

Using a 5 percent exclusion level (the probability of rejecting the null hypothesis, H_o) with three degrees of freedom the chi-square test statistic is:

$X^2_{0.05,\,3} = 7.815$

H_o is rejected if X^2 for our data set is greater than 7.815. For the overall evaluation of the data the chi-square test statistic is:

$X^2 = 24.112$

This value is greater than 7.815 and H_o is rejected. Therefore, the alternative hypothesis H_a is accepted and a conclusion can be drawn that the detectability of a defect is dependent on the size of the defect. Also to support this hypothesis are the p-values from Minitab and SAS programs. The p-values for the overall model are:

Minitab = 0.000 (probability is very small – strong evidence of association)
SAS = 0.001 (again probability is very small – strong evidence of association)

Hence, for any p-value less than the 5% exclusion level there is evidence of a relationship between size and success of detection.

Logistic Regression Test

The logistic procedure is a new statistic to study the relationship between two or more variables (Trindade, 1998). Again, our variables are success of detecting a defect and the size of a particular defect. The logistic regression test uses a Hosmer and Lemeshow Goodness-of-Fit test to tell if the logistic regression is a good model for our data. In the goodness-of-fit test, the expected frequencies should all be above five, with the exclusion of a few samples below. The p-value must also be non-significant, which is greater than the 5 percent exclusion level. By testing each evaluation separately, the expected frequencies fall below five and therefore the logistic model is not a good fit. This outcome is a result of using a sample size of 44. With the overall model, the expected frequencies are above five with the exception of three samples. This level of anomalous results is acceptable (Smith, 1998). Table 4.2 shows the expected frequencies and a non-significant p-value.

Table 4.2: Hosmer and Lemeshow Goodness of Fit Table for Overall Evaluations

Success = Detected		**Success = Not Detected**	
Observed	Expected	Observed	Expected
15	16.61	15	13.39
16	17.44	14	12.56
15	14.36	9	9.64
23	19.33	7	10.67
14	12.41	4	5.59
13	17.45	11	6.55
19	18.77	5	5.23
21	19.83	3	4.17
23	21.00	1	3.00
33	34.82	3	1.18

Goodness-of-fit Statistic = 12.284 with 8 DF
P-value = 0.1390

Using the logistic regression test it is possible to calculate the odds ratio for a 10 percent increase in size with a 95 percent confidence interval (SAS, 1991).

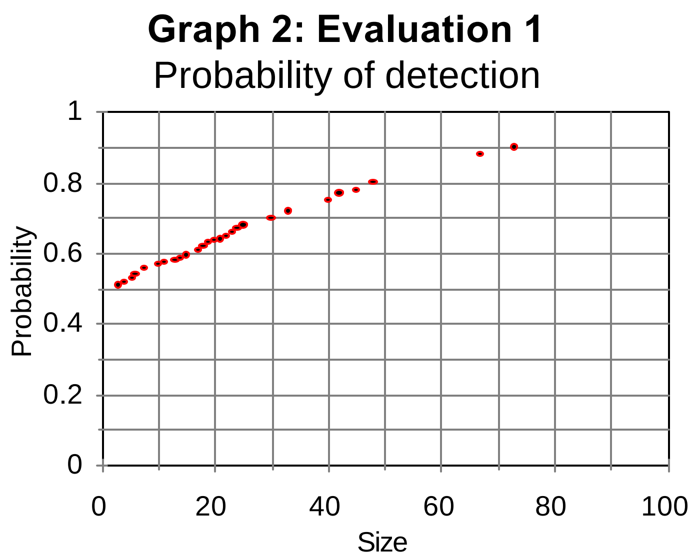
Table 4.3: Odds Ratios and 95% Confidence Interval for Overall Evaluations

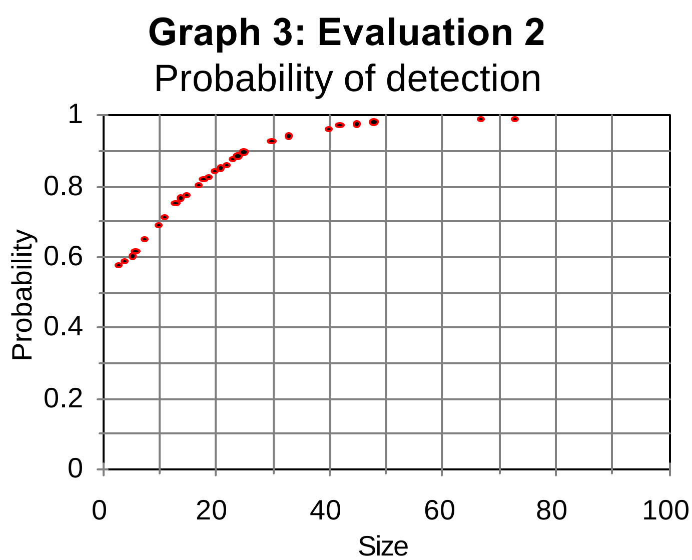
			Wald Confidence Limits	
Variable	Unit	Odds Ratio	Lower	Upper
Size	10.0	2.023	1.469	2.786

As long as the confidence interval does not include one and the odds ratio is above one then this test gives support to this analysis. For the overall model the odds of detecting a defect are twice as high for a defect of size 10 percent or higher as compared to one of a size lower than 10 percent.

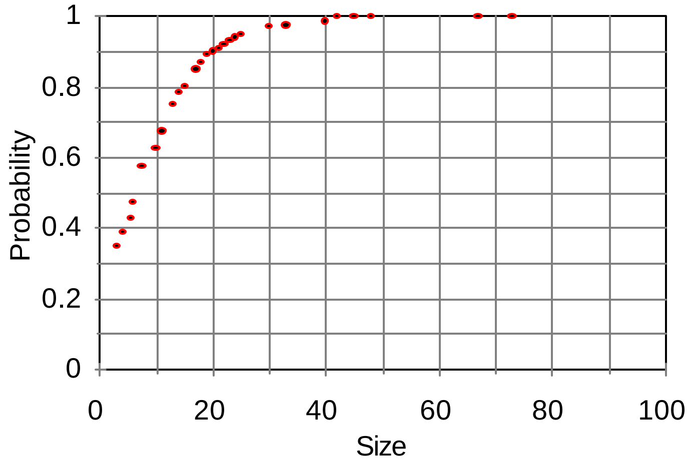
Probability of Detection Plots

The probability of detection plots were obtained using the SAS program. These show the probability of an inspection vehicle detecting a defect of a certain size. Graphs 2 through 7 show the probability of detection for evaluations 1 through 6. These evaluations have a sample size 44. Graph 8 provides the overall probability of detection and it uses the results from evaluations 1 through 6 for a sample size of 264. All seven graphs show that as a defect increases with size the probability of detecting that defect increases.

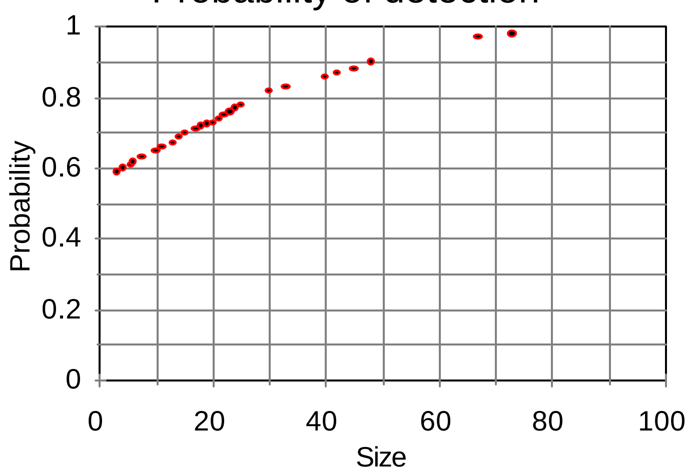




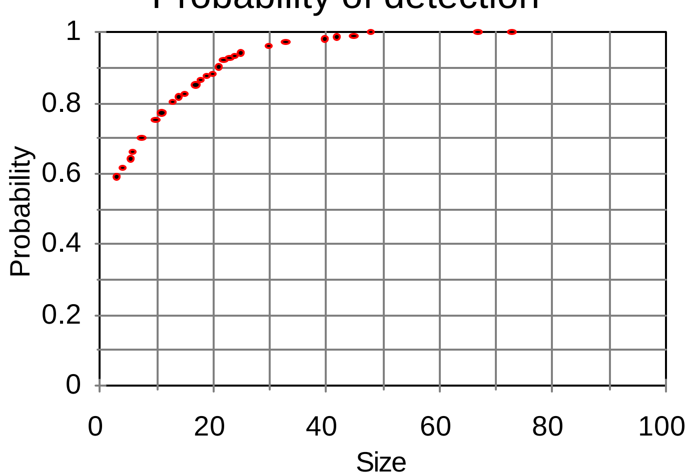
Graph 4: Evaluation 3
Probability of detection

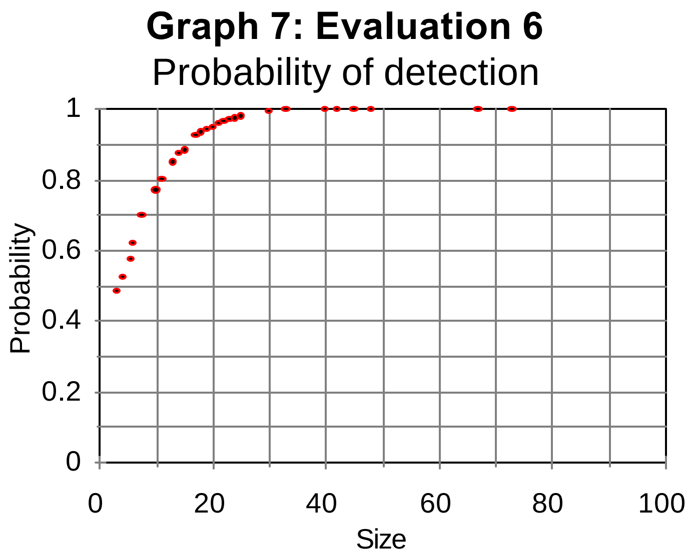


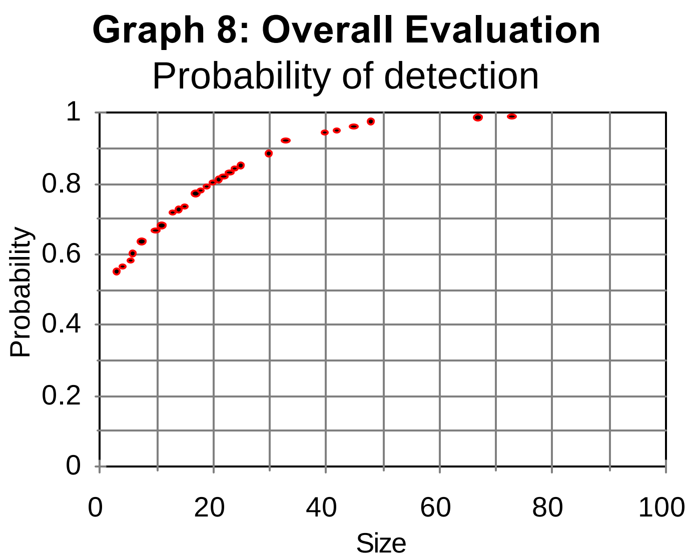
Graph 5: Evaluation 4
Probability of detection



Graph 6: Evaluation 5
Probability of detection







Graph 8 shows the overall samples evaluated together. The overall evaluation of the inspection vehicles provides the largest sample size and produces the most important results. Rail inspection reliability is a combination of the results of all evaluations. Hence, the overall probability of detection is comparable to the reliability of the inspection system. The overall probability of detection also should be compared to the inspection reliability. The inspection reliability standards are in the 1997 American Railway Engineering Association (AREA) Manual of Recommended Practices, Chapter 2. These are recommendations, not standards.

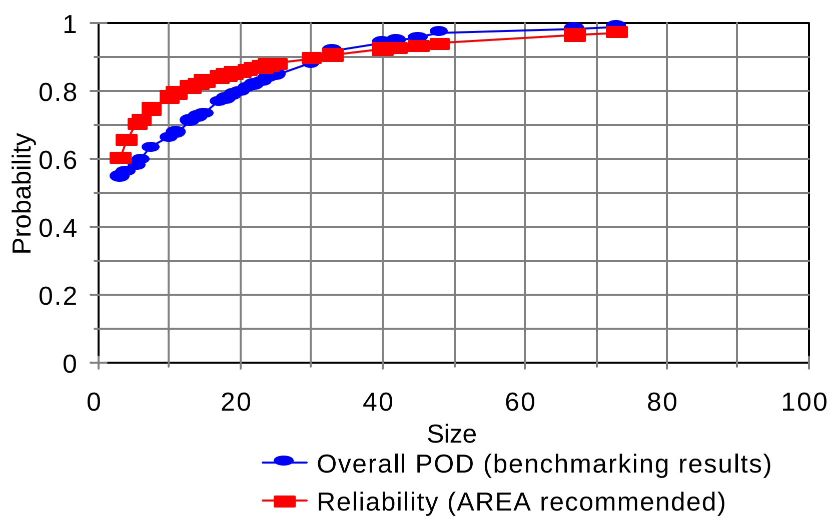
5. CONCLUSIONS

If inspection frequency has not been adjusted to compensate for increased defect occurrence, the deterioration of rail caused by aging gradually reduces the factor of safety in an otherwise conservative design. The inspection reliability is a measure of determining the level of risk. The inspection reliability is related to defect size and the ability to detect defects. The rail industry must constantly rehabilitate and maintain the rail system to keep it safe for its customers and the public. Therefore, a reliable and accurate inspection system is necessary for finding defects in the rail. Accuracy of the inspection system refers to size of the defect that can be detected and reliability refers to probability of detecting a defect of a given size. The comparison of the overall probability of detection plot from benchmarking efforts at TTC (Graph 8) and the AREA reliability recommendations is shown in Graph 5.1 (TTCI TD 1997).

Inspection Reliability = IR = $[0.217 \ln(\% \text{ defect size})]^n$

Where n = 0.35 for detail fractures (industry standard). The inspection reliability describes industry assumed level of performance based on previous experimental work and provides a degree of confidence with the inspection technology.

Graph 5.1: Inspection Reliability and Probability of Detection Comparison



This correlation shows the probability of detection curve and the AREA recommended inspection reliability curve for detail fractures. The reliability curve and the overall evaluation probability of detection curve increase with defect size. The inspection reliability curve supercedes the probability of detection curve for defects smaller than 30 percent. From this comparison, the overall probability of detection results for the RDTF and the minimum inspection reliability, one can conclude the ultrasonic inspection technology has a lower probability of detecting smaller defects than the expected AREA recommended industry reliability. However, the AREA manual only is a recommended guideline not a required specification.

REFERENCES

Davis, D. D. (1997). "Rail Defects and their Effects on Service Reliability," AAR. From Expanded Workshop: Rail Defect and Broken Rail Detection, TTCI, July 23.

Clarke, R. (1968). "Rail Testing – Where we go from here," Sperry Rail Service, From the Bulletin of the National Railway Historical Society, Vol. 33 No. 6.

Franke, M. W. (1997). "Managing our Key Railroad Asset," BNSF, From Expanded Workshop: Rail Defect and Broken Rail Detection, TTCI, July 22-23.

Garcia, G. A. (1998). "Methods of Inspection," TTCI, June 9.

Krautkramer, J. and Krautkramer, H. (1983). "Ultrasonic Testing of Materials," Springer-Verlag Berlin Heidelberg New York, Third Edition, pages 30-31.

Ryan, T., Juiner, B., Ryan, B. (1976). "Minitab Student Handbook," Duxbury Press, North Scituate, Massachusets.

SAS/QC Software (1991). SAS Institute Inc, NC, Version 6.12 First Edition.

Smith, J. (1998). Statistics Masters candidate, Colorado State University, Fort Collins, Colorado.

Trindade, A. (1998). Statistics Ph.D. candidate, Colorado State University, Fort Collins, Colorado.

Zarembski, A. M. (1997). "Review of Rail Testing Technologies," Zeta – Tech Assoc. Inc., From Expanded Workshop: Rail Defect and Broken Rail Detection, TTCI, July 22-23.