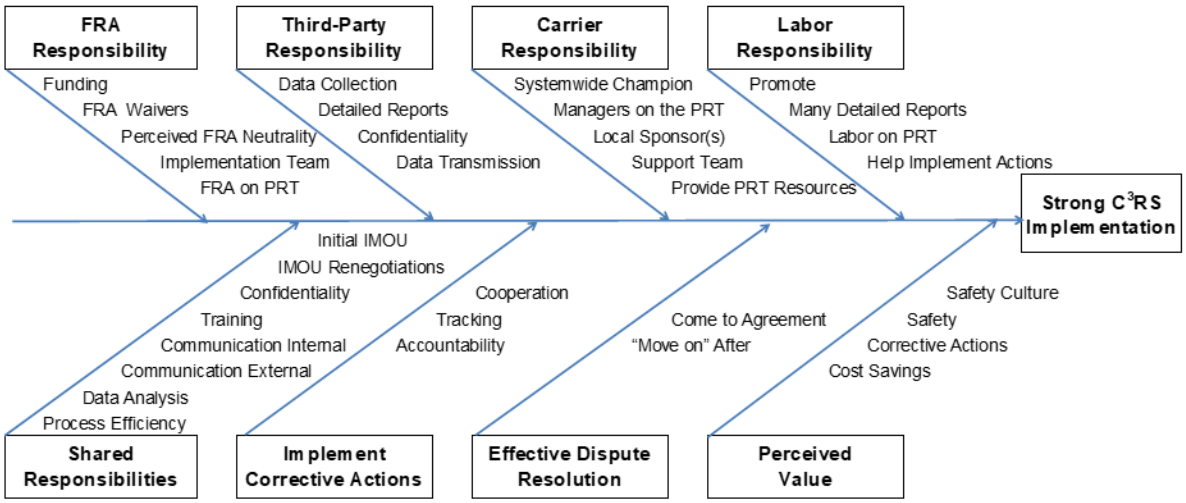
U.S. Department of Transportation

Federal Railroad Administration

Confidential Close Call Reporting System (C³RS) Lessons Learned Evaluation – Final Report

Office of Research,
Development
and Technology
Washington, DC 20590



Implementation Factors for FRA's Confidential Close Call Reporting System

DOT/FRA/ORD-19/01

Final Report
February 2019

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 2019	Technical Report 2006–August 2017

4. TITLE AND SUBTITLE
Confidential Close Call Reporting System (C^3RS) Lessons Learned Evaluation – Final Report

5. FUNDING NUMBERS
RR24A2-PY430
RR24A2-QY430

6. AUTHOR(S)
Joyce M. Ranney (Volpe Center), Melinda Davey (Jacobs), Jonathan Morell (Syntek Technologies Inc.), Michael Zuschlag (Volpe Center), Starr Kidda (Federal Railroad Administration)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
John A. Volpe National Transportation Systems Center
55 Broadway
Cambridge, MA 02142-1093

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, Development and Technology
Washington, DC 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DOT/FRA/RPD-19/01

11. SUPPLEMENTARY NOTES
COR: Dr. Maryam Allahyar

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the FRA website.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
The Federal Railroad Administration (FRA) established the Confidential Close Call Reporting System (C^3RS) program through which "close calls" can be reported confidentially and addressed by Peer Review Teams (PRT) containing representatives from labor, management, and FRA. Close call programs have shown to work in other industries, but their effectiveness in the railroad industry was an open question. To address this question, FRA implemented a rigorous lessons learned evaluation over a 12 year duration across four demonstration pilot sites. The C^3RS evaluation is designed to answer three major questions: (1) What conditions are necessary to implement C^3RS as planned in a demonstration? (2) What is the impact of C^3RS on safety and safety culture? (3) What factors help to sustain C^3RS long-term, beyond the demonstration? This report describes the final answers to the evaluation questions. In summary, implementing C^3RS as planned is possible within transportation departments in the railroad industry. Bottom-line impacts were achieved in the presence of C^3RS in areas such as reduced derailments (3 sites), injuries (1 site), discipline hearings (two sites) and improved safety culture (4 sites). C^3RS can be sustainable in railroad transportation when enlisting the support of local labor, management, national labor, and FRA.

14. SUBJECT TERMS
Close calls, continuous improvement, employee involvement, freight rail, human factors, passenger rail, risk reduction, safety culture, Confidential Close Call Reporting System, C^3RS

15. NUMBER OF PAGES
219

16. PRICE CODE

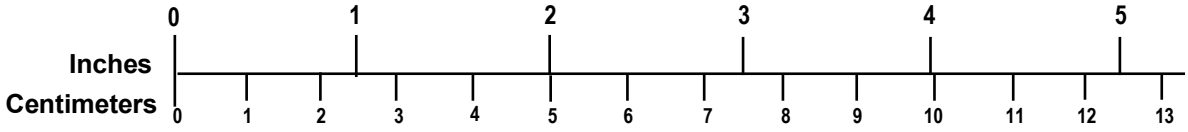
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500
Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

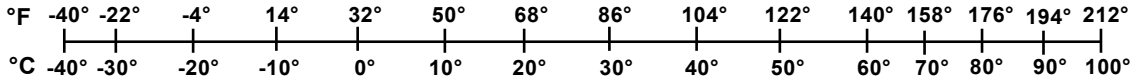
METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC			METRIC TO ENGLISH		
LENGTH (APPROXIMATE)			**LENGTH (APPROXIMATE)**		
1 inch (in)	=	2.5 centimeters (cm)	1 millimeter (mm)	=	0.04 inch (in)
1 foot (ft)	=	30 centimeters (cm)	1 centimeter (cm)	=	0.4 inch (in)
1 yard (yd)	=	0.9 meter (m)	1 meter (m)	=	3.3 feet (ft)
1 mile (mi)	=	1.6 kilometers (km)	1 meter (m)	=	1.1 yards (yd)
			1 kilometer (km)	=	0.6 mile (mi)
AREA (APPROXIMATE)			**AREA (APPROXIMATE)**		
1 square inch (sq in, in^2)	=	6.5 square centimeters (cm^2)	1 square centimeter (cm^2)	=	0.16 square inch (sq in, in^2)
1 square foot (sq ft, ft^2)	=	0.09 square meter (m^2)	1 square meter (m^2)	=	1.2 square yards (sq yd, yd^2)
1 square yard (sq yd, yd^2)	=	0.8 square meter (m^2)	1 square kilometer (km^2)	=	0.4 square mile (sq mi, mi^2)
1 square mile (sq mi, mi^2)	=	2.6 square kilometers (km^2)	10,000 square meters (m^2)	=	1 hectare (ha) = 2.5 acres
1 acre = 0.4 hectare (he)	=	4,000 square meters (m^2)			
MASS - WEIGHT (APPROXIMATE)			**MASS - WEIGHT (APPROXIMATE)**		
1 ounce (oz)	=	28 grams (gm)	1 gram (gm)	=	0.036 ounce (oz)
1 pound (lb)	=	0.45 kilogram (kg)	1 kilogram (kg)	=	2.2 pounds (lb)
1 short ton = 2,000 pounds (lb)	=	0.9 tonne (t)	1 tonne (t)	=	1,000 kilograms (kg) = 1.1 short tons
VOLUME (APPROXIMATE)			**VOLUME (APPROXIMATE)**		
1 teaspoon (tsp)	=	5 milliliters (ml)	1 milliliter (ml)	=	0.03 fluid ounce (fl oz)
1 tablespoon (tbsp)	=	15 milliliters (ml)	1 liter (l)	=	2.1 pints (pt)
1 fluid ounce (fl oz)	=	30 milliliters (ml)	1 liter (l)	=	1.06 quarts (qt)
1 cup (c)	=	0.24 liter (l)	1 liter (l)	=	0.26 gallon (gal)
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 meter (m^3)	=	36 cubic feet (cu ft, ft^3)
1 cubic yard (cu yd, yd^3)	=	0.76 cubic meter (m^3)	1 cubic meter (m^3)	=	1.3 cubic yards (cu yd, yd^3)
TEMPERATURE (EXACT)			**TEMPERATURE (EXACT)**		
[(x-32)(5/9)] °F	=	y °C	[(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

Acknowledgments

This study would not have been possible without the cooperation of many people. We would like to thank the National C^{3}RS Steering Committee and those who participated in the Federal Railroad Administration's (FRA)-sponsored human factors workshop, *Improving Railroad Safety through Understanding Close Calls*, in 2003, which led to the start of the C^{3}RS demonstration.[1]

Thank you to those who participated in the Lessons Learned interviews and stakeholder reviews of findings and deliverables: labor and managers from the demonstration pilot site railroads, FRA (including the inspectors who sit on the C^{3}RS Peer Review Teams, Deputy Regional Administrations, Office of Railroad Policy and Development and Office of Railroad Safety), the C^{3}RS Implementation Team, BTS, and the National Aeronautics and Space Administration (NASA).

Thank you to Michael Coplen, David Moore, Linda Connell, Robert Castiglione, and Brian Reilly for your reviews and support.

[1] *Improving Railroad Safety Through Understanding Close Calls*. (FRA) Workshop Proceedings. (Accessed online July 9, 2015.)

Contents

Illustrations

Tables

Executive Summary

Historically, concern of litigation and a blame-based safety culture in the railroad industry have prevented free and open discussion of dangerous conditions and how they might be eliminated. As a result, accidents occur that might well have been prevented. In recognition of this problem, the Federal Railroad Administration (FRA) established a program through which "close calls" could be reported confidentially. This innovative program, known as the Confidential Close Call Reporting System (C^3RS), defined a close call as "an opportunity to improve safety practices in a situation or incident that has a potential for more serious consequences." Peer Review Teams (PRT) were created to facilitate open and honest communication among representatives of management, labor, and FRA regional inspectors as they analyzed the close call data.

Close call programs have been shown to work in other industries, but their effectiveness in the railroad industry was an open question. To address this, FRA implemented demonstration pilots at Union Pacific (UP) North Platte Service Area, Canadian Pacific (CP) Chicago Area, New Jersey Transit (NJT), and Amtrak. The demonstration pilots were conducted by the Volpe National Transportation Systems Center with support from Jacobs Engineering.

FRA sponsored a rigorous quasi-experimental, mixed-methods-design evaluation that began in 2006 and ended in August 2017. The evaluation sought to answer three questions.

1. What conditions are necessary to implement C^3RS as planned in the railroad industry?
2. What is the impact of C^3RS on safety and safety culture?
3. What factors help to sustain C^3RS long-term, beyond the demonstration?

The research design combined measurements over time and across the four different demonstration pilot railroads (i.e., two freight and two passengers). Data included interviews with participants and stakeholders, trend analysis of close call reports and corrective actions, validated scales of safety culture, and "bottom-line" impacts to safety. The evaluation began at the same time as the C^3RS demonstration implementations, thus affording a rich opportunity to understand C^3RS.

The evaluation collected data from the four demonstration sites for a period of 5 years each during baseline, midterm, and final phases. Due to the non-simultaneous initiation of the demonstration pilots at each railroad, the actual data collection spread over a 10-year period. This report includes the final findings from the evaluation.

The evaluation revealed that implementing C^3RS as planned is possible within transportation departments in the U.S. railroad industry. While the demonstrations had varying degrees of implementation, each of them performed sufficient planned activities to justify an evaluation of their impacts.

Improvements in safety and safety culture were achieved in the presence of C^3RS: derailments (three sites), injuries (one site), discipline hearings (two sites), and safety culture (four sites). The Evaluation Team assessed multiple types of data to determine appropriate impact metrics at each site. At three sites, derailments were reduced between 20–41 percent in the presence of C^3RS implementation (Figure 1).

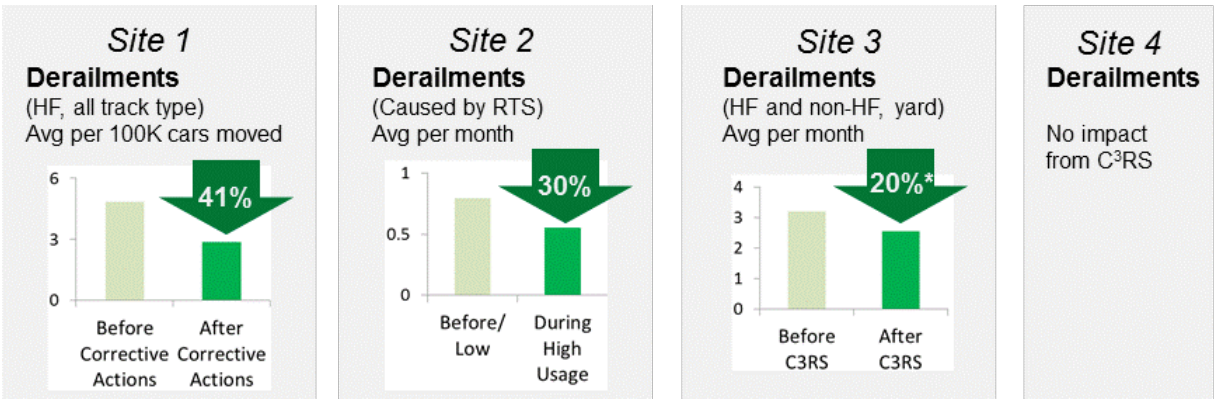


Figure 1. Summary of Impacts on Derailments[2]

Achieving improvement in safety through C^{3}RS requires commitment and active participation from labor, management, and FRA. Common aspects of favorable implementation were: 1) support for C^{3}RS from a champion with system wide influence; 2) enthusiastic involvement from labor in helping carriers to implement corrective actions; 3) constructive dispute resolution; and 4) agreement between labor and management that C^{3}RS had worthwhile impact.

Successful implementation by carrier management and labor was necessary but not sufficient for long-term sustainability after the demonstration. For FRA's model of C^{3}RS to be sustainable at a site in the long run, there must be consensus as to how C^{3}RS will be structured and how it will operate. The agreement has to work in multiple ways. Management and labor within the site must agree on how internal C^{3}RS operations will take place. The site and FRA must agree on reporting procedures that correspond to discipline waivers. Labor at the site, carrier management, and FRA must maintain the support of labor at the national level. All parties must accept the services provided by National Aeronautics and Space Administration (NASA) as the C^{3}RS Third-Party data collector.

C^{3}RS can be beneficial and sustainable in the railroad industry with both good implementation by individual carriers and continued support from FRA and national labor. FRA has taken steps to support sustainability in the railroad industry. Those steps have included continued funding for the C^{3}RS Third-Party to collect close call reports, the allocation of dedicated staff, and the transfer of C^{3}RS from FRA's Office of Research, Development and Technology (RD&T) to the Office of Railroad Safety's (RRS) Human Performance Division (HPD). As a result of these efforts, C^{3}RS now has eight carriers participating (i.e., two original demonstration sites and six new carriers). Also, the number of labor unions involved has grown from two in the original pilots (i.e., BLET and SMART-TD) to 15. C^{3}RS began with just transportation and expanded to include engineering and mechanical crafts. Despite this progress, the sustainability is limited to passenger railroads. No dedicated freight railroad is part of the FRA's C^{3}RS program.

To replicate C^{3}RS's implementation, impact, and sustainability successes, a railroad must engage in activities that were not all known at the program start. Railroad management, labor, and FRA can use the following lessons learned to achieve effective C^{3}RS processes at other sites.

[2] A decrease marked with an * indicates that it is not statistically significant.

Implementation:

- Detailed reports are more important than a large number of reports for guiding the development of corrective actions.
- Labor participation in corrective action implementation is essential for effective C^3RS operation within a site.
- Management needs to be effective in providing the PRTs with the expertise needed to build a business case for corrective actions.
- Managers should resolve disputes over the scope of C^3RS discipline protection in a way that encourages future reporting.
- Personnel transitions on C^3RS teams can cause downtime and negatively affect bottom-line impact.
- Sites can share knowledge about non-proprietary process improvements and corrective actions. Such sharing increases the overall benefit of C^3RS in the railroad industry.
- Sites have limited capacity to implement corrective actions; therefore, guidance on priorities is needed to help PRTs prioritize their analysis efforts.
- Participation in PRT meetings by FRA regional personnel provides significant value, but may not be required at all meetings.
- Reasons why railroads find it difficult to track corrective actions should be addressed to improve understanding of accomplishments and provide guidance for future activities.

Impact:

- Close call data are useful. Because railroads can discover safety problems that they did not know about, they should encourage reporting.
- Carriers should monitor corrective actions to determine if they are effectively solving problems and improving safety.
- Derailment close call reports seem to be useful in reducing derailments and encouragement could be helpful.
- Safety culture improvements in supervisor-employee relationships can be attained. Supervisors should embrace this willingness to communicate and have productive, blame-free conversations about safety with their employees.

Sustainability:

- FRA should determine ways to measure impacts in the future to ensure continued justification and funding for the program.
- Sustainability is strengthened by support from stakeholders external to a carrier. The suitability of the C^3RS model should be continually reconsidered with respect to changes in technology, the needs of stakeholders, and conditions in the industry.
- FRA should consider new approaches to reach out to Class 1 railroads.
- Sustainability is strengthened when success stories are shared.

1. Introduction

1.1 Background

Since 2003, the Federal Railroad Administration (FRA) has implemented a demonstration project entitled the Confidential Close Call Reporting System (C^3RS) through a C^3RS Implementation Team. A separate C^3RS Evaluation Team initiated their study in 2005, whose role was to provide a scientifically credible evaluation of the C^3RS demonstration project. The Evaluation Team prepared this final report containing the findings for the multi-year, multiple railroad C^3RS Lessons Learned Evaluation, including data collection and analysis methods.

1.1.1 Context: Challenges to Improving Safety

1.1.1.1 Challenges in Rail

In 2002, FRA's Office of Research, Development and Technology's (RD&T) Human Factors Division (HF) observed that there had not been any major improvements in human factors-caused reportable incident levels in the preceding decade (for example, see Figure 2). The FRA accident database provided information on what happened (the types of incidents) but not why incidents happened. The RD&T HF theorized that collecting and studying close call information directly from people involved in events could provide valuable information on safety hazards and help railroads reduce incident levels. The C^3RS Implementation Team defined a close call as "an opportunity to improve safety practices in a situation or incident that has a potential for more serious consequences." To manage safety proactively, close calls could be collected and analyzed to identify patterns related to failures or weaknesses in the system.

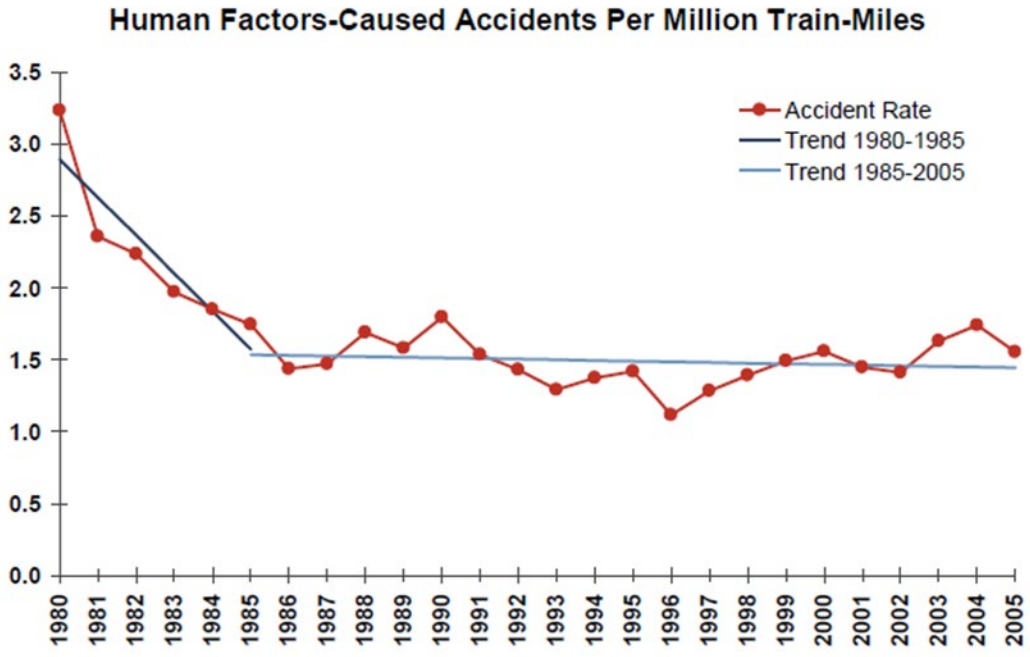


Figure 2. Historical Human Factors Accidents Before C^3RS Began

The C^{3}RS Implementation Team identified the following potential benefits from collecting close call data, as opposed to having knowledge of only reportable events:

- *Close calls can show where current weaknesses exist in the safety system, including those that were previously unknown*. Hidden safety risks may exist for a while before an incident occurs (Reason, 1998).
- *Reported close calls can be used to monitor changes in safety over time*. Since close calls may be more plentiful than reportable incidents, they may provide more information for assessing increases or decreases in safety trends.
- *Close call data can reveal useful information.* Close calls can reveal better information about human performance and interactions between people and technology so organizations can proactively address the correct problems and prevent incidents.

However, RD&T HF knew that achieving improved incident levels using close call information would take significant cooperation between labor and management. Therefore, any close call program in the rail industry would require a concerted focus on improving labor–management relations. In the railroad industry, this is often referred to as safety culture.

Historically, the railroad industry has had contentious labor-management relations. Management style was focused on a command-and-control structure and disciplining employees for safety rule infractions causing employees to be concerned regarding disclosures about safety (Coplen, 1999). Adding to the conflict is the controversy over the hundred-year-old Federal Employer's Liability Act (FELA).[3] [4] In 1908, Congress determined that the railroads could not use the no-fault Workman's Compensation insurance for personal injuries because it was limited to Federal workers in "hazardous life-saving work." FELA was enacted, and it was a blame-based liability system for personal injuries. FELA allowed railroad employees to sue their employer for damages if they were injured. FELA also allowed management to only pay compensation, if the injured railroader proved the railroad was legally negligent (TRB, 1994). The amount of compensation was dependent on how much of the blame for the injury fell on the employee vs. the employer. This created a strong financial incentive for each side to directly blame the other, inhibiting the free and open discussion needed to understand why an incident happened (Zuschlag, 2012).

1.1.1.2 Challenges in Industries Other than Rail

Airline, chemical processing, nuclear, and transportation industries have also been concerned with unreported risks, which led them to establish programs to collect close call information. Across these industries, the following issues were considered in setting up a close call program (Morell, 2006):

- Incentives for reporting and protections for reporters
- Identification of who reviews the reports and system ownership
- Problem-solving analysis methods

---

[3] 45 U.S.C. § 51 et seq. (1908).

[4] National Research Council. Compensating Injured Railroad Workers Under the Federal Employer's Liability Act: Special Report 241. Washington, DC: The National Academies Press, 1994.

- Characteristics of problem-solving groups
- How to implement change
- Consequences in terms of safety and safety culture
- Sustainment and evolution of close call systems over time

Within the aviation industry alone, systems have been developed and deployed in the U.S., Canada, Australia, U.K., New Zealand, Germany, Japan, Korea, and Taiwan (Sullivan, 2001). The U.S. aviation industry has used the Aviation Safety Reporting System (ASRS) since 1976.[5] The National Aeronautics and Space Administration (NASA) administers the program and collects aviation safety incident/situation reports voluntarily submitted by pilots, controllers, and others. Reporters are granted protection from discipline in return for providing the close call data. ASRS staff analyzes the data and feeds it into a public repository. ASRS staff educates the airline industry about system deficiencies and safety risks through a newsletter, journal, and research studies. As of 2013, ASRS received over 1.1 million reports.

[5]NASA's Aviation Safety Reporting System (ASRS). Accessed online: June 23, 2015

Definitions

Close Call Event – a specific situation or event, that has a potential for more serious consequences, that was observed by a railroad employee.

Known Close Call Event – an event that is below the FRA reporting threshold for operating rules and does not involve an injury, but would require managerial notification if discipline protection was sought through C^{3}RS. To facilitate analysis of such events, employees provide notification of the event to management without undue delay in addition to the C^{3}RS report.

C^{3}RS Report – created when an individual railroad employee observes a close call event and submits a written report to the Third-Party, and it is accepted under the criteria laid out in that railroad's IMOU.

C^{3}RS-Report-Record – created when a given C^{3}RS Report is de-identified by the C^{3}RS Third-Party. It can originate from a single employee's C^{3}RS report or several employees' reports about the same close call event. The Third-Party consolidates the information from the original report(s), conducts follow-up interviews with the reporter(s) as available, removes identifying information, and creates a written C^{3}RS-report-record. Then the Third-Party sends report-records to the PRT for analysis.

C^{3}RS-Analyzed-Case – created when a report-record is analyzed by the PRT. It contains the results from the PRT's analysis.

Corrective Actions – actions that a railroad can take to mitigate a safety risk. The PRT creates recommendations for corrective actions based on C^{3}RS-analyzed-cases. Corrective actions can be specific to a particular location, or applicable to a wider geographical area and/or across organizational boundaries.

See the Glossary for more.

1.1.2 Solution: FRA Confidential Close Call Reporting System

1.1.2.1 Description of C^{3}RS Process

In 2003, FRA set up its own reporting system entitled the Confidential Close Call Reporting System (C^{3}RS) as a solution to the challenges described above, to proactively manage safety by systemically studying close calls. Figure 3 shows the C^{3}RS process for a committed demonstration site. The activities occurring at FRA and the industry were not included in the figure. Definitions of terms used in this report are included to the Right, and also appear in the Glossary. Similar to ASRS, C^{3}RS provides a quid pro quo where reporters are granted indemnity in return for supplying close call data to the program.

The process began when a worker observed a **close call event** and submitted a **C^{3}RS report** to the **C^{3}RS Third-Party**, thereby assuring the worker's confidentiality. Both NASA and the Bureau of Transportation Statistics (BTS) served as the Third-Party during the C^{3}RS demonstration. In some situations, an employee may be allowed to report a **known close call event**, e.g., a run-through switch,[6] and receive discipline protection in exchange for their report.

The Third-Party determined if they had received multiple reports about the same close call event. Then the Third-Party consolidated the information from the original report(s), conducted follow-up interviews with the reporter(s) as available, removed identifying information, and created a written **C^{3}RS-Report-Record**. The Third-Party regularly provided the de-identified C^{3}RS-report-records to a **Peer Review Team (PRT)** at the applicable railroad.

The purpose of the PRT was to analyze the report-records and make recommendations to management. The PRT typically consisted of representatives from labor, management, and FRA regional inspectors, all of whom had been trained in a problem-solving methodology known as Multiple Cause Incident Analysis (MCIA). A **C^{3}RS-Analyzed-Case** contained the report-record and the results from the PRT's MCIA. The PRT then created recommendations for **Corrective Actions** based on one or more C^{3}RS-analyzed-cases to improve safety. The PRT sent the recommendations to a **Support Team**, comprised of middle and senior management, which was typically established to be responsible for the review and implementation of corrective actions. The railroad monitored corrective actions implemented and shared information about them with the workforce. Finally, after enough railroads had joined C^{3}RS, the Third-Party analyzed reporting trends across carriers and shared results with FRA and the industry.[7]

---

[6] A run-through switch occurs when train goes through a misaligned switch and damages it.

[7] For the purpose of this report, the term "trends" refers to an analysis of the patterns of the information in the close call reports that were submitted and accepted. It does not refer to the population of all close call events in the four sites that occurred during the evaluation period.

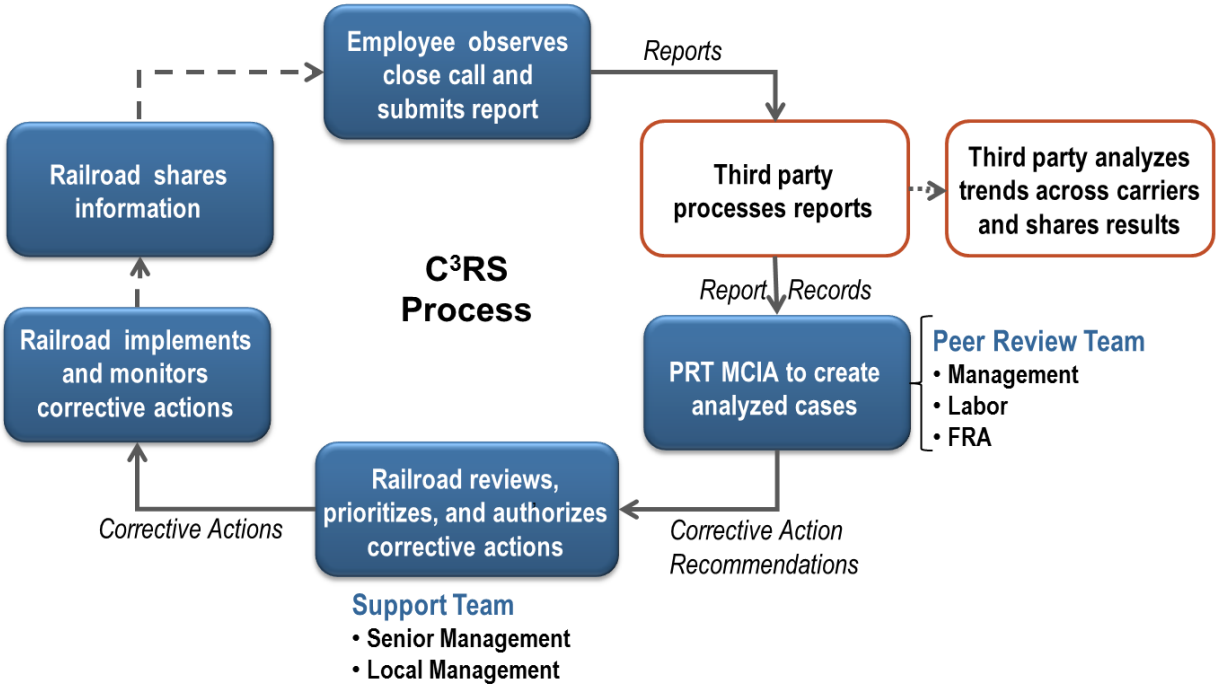


Figure 3. C³RS Process

1.1.2.2 Key Features of C³RS in the Railroad Industry

What makes C³RS unique in the railroad industry is its use of mechanisms borrowed from aviation and other close call programs that allows people to report close calls and minor violations without concern of discipline, thus providing information to railroad labor, management, and FRA that they would not otherwise have known. The information reported is analyzed to find causes, and corrective actions are developed to address the unsafe situation. Seven key mechanisms exist:

1. **Use of a Third-Party**. To facilitate a free and open discussion of close calls, a trusted mechanism exists through which workers can report close call events so that (a) the reporter's identity is protected, but that also: (b) provides enough detail for a problem-solving team, the PRT, to be able to analyze the situation and understand it well enough to formulate a corrective action. The Third-Party shares no report-records or other information with the PRT or the industry that has not been de-identified of all personally protected information (name, etc.). Moreover, while carriers know that report-records came from their workforce, neither FRA nor the industry knows the name of the railroad for any trend information provided by the Third-Party.
2. **Confidentiality and Protection**. The Third-Party de-identifies the C³RS-report-records before providing them to the railroads. Reporters are given protection from discipline in the case of a "known close call event" as determined by their implemented memorandum of understanding (IMOU), in return for the information provided in their report, in which case their identity would already be known to the railroad. Participating railroads are also given confidentiality and protection from FRA sanctions for reported events. Confidentiality is of paramount importance for a railroad close call program because of a history of distrust and litigation in the industry.

3. **Use of a Railroad Peer Review Team Containing Three Stakeholders**. PRT Team membership consists of carrier management, carrier labor unions, and local FRA inspectors. In order to partner effectively, the stakeholders are trained to work as a group to apply the MCIA as a method for finding causes and determining corrective actions.
4. **FRA's Involvement.** Besides their involvement on the PRT, FRA also grants waivers that legally permit participating railroads to refrain from taking disciplinary action in situations where discipline would normally be required. The waivers concern Title 49 Code of Federal Regulations (CFR) Part 240, Qualification and Certification of Locomotive Engineers, a provision that requires engineers to lose their certification to operate a locomotive if they violate specific operating rules. The waiver protects railroads from fines imposed by FRA in the event that an employee was not disciplined. Another waiver is for Part 242, which requires conductor certification. (Part 242 took effect several years after C^3RS first began, at which point waivers began to be needed to cover conductors.)
5. **Use of an IMOU to Document Agreements Among the Stakeholders.** A C^3RS IMOU is negotiated among labor, management, and the FRA for a given railroad. The IMOU clearly articulates close call events that are and are not included within the program's scope, the requirements for confidentiality, and the obligations and commitments of all parties to the agreement. Examples of topics that are covered are: known close call events and conditions under which discipline can be avoided, how many crew members need to report for a report to be valid, and how disputes will be resolved. Instructions on which "known close call events" are eligible for protection from discipline are in Section 6.4 of the IMOU, so C^3RS participants often refer to discipline protection as "6.4." The IMOU was expected to be a flexible document that could be changed as the needs of the internal and external stakeholders changed over time.
6. **Use of a Support Team to Evaluate and Implement Corrective Actions.** A designated Support Team comprised of middle and senior managers is responsible for reviewing, approving, and implementing corrective action. PRT members can sometimes work on a corrective action themselves, or local managers can effect change on their own authority. Other times senior management approval is needed, as might be the case if substantial funds were required, or a company-wide policy had to be revised. The Support Team is comprised of senior managers who have the organizational authority to review corrective actions, decide which ones to implement, and oversee the implementation of corrective actions. The Support Team communicates its assessment of the worthiness of all corrective actions submitted and their plans for implementation if needed to the PRT. Some railroads, especially smaller ones, may decide to assign a dedicated senior manager instead of a formal Support Team.
7. **Railroad's Agreement for the Third-Party to Share the Aggregated Trends from Reporting Data.** The railroads in C^3RS agree that aggregated close call data, once de-identified, can be shared with FRA and the industry. It is the responsibility of the Third-Party to publish newsletters with aggregated information and alerts for the industry. Only the FRA inspector who is a member of the PRT learns directly about report-records (de-identified). As a PRT member, he/she is included in their analyses but not allowed to discuss specific details outside the PRT.

1.2 Objectives

This report evaluates the FRA-sponsored C^{3}RS demonstration project. FRA's RD&T) Human Factors (HF) Division decided that an external evaluation was needed because C^{3}RS was both ambitious and risky. It was ambitious because its goal was to try a new method for the U.S. railroad industry to improve safety. RD&T HF sponsored a demonstration project comprised of the first four railroads to join C^{3}RS and planned for multiyear participation, which was risky because RD&T HF committed to spending a significant amount of money, time, and political capital on an experiment that could fail in numerous ways. RD&T HF anticipated that there would be champions as well as detractors. To ensure an unbiased evaluation was conducted as part of the demonstration, they included an external evaluation.

The objectives of the evaluation were to answer three questions:

1. What conditions are necessary to implement C^{3}RS as planned in the railroad industry?
2. What is the impact of C^{3}RS on safety and safety culture?
3. What factors help to sustain C^{3}RS long-term, beyond the demonstration?

1.3 Overall Approach

The evaluation has a quasi-experimental, mixed-methods design that was both formative and summative. It was formative in the sense that it was intended to inform implementation and summative in that it planned to provide conclusions about C^{3}RS's impacts.

The research design combined measurements over time and across the four different demonstration pilot railroads (i.e., two freight and two passengers). Data included interviews with participants and stakeholders, trend analysis of close call reports and corrective actions, validated scales of safety culture, and "bottom-line" impacts to safety. The evaluation began at the same time as the C^{3}RS demonstration implementations, thus affording a rich opportunity to understand C^{3}RS.

The evaluation collected data from the four demonstration sites for a period of 5 years each during baseline, midterm, and final phases. Due to the non-simultaneous initiation of the demonstration pilots at each railroad, the actual data collection spread over a 10-year period. This report includes the final findings from the evaluation.

1.4 Scope

This report is limited to evaluation results from the four demonstration sites and the interpretation of those results into findings by the Evaluation Team. The four sites were: Union Pacific Railroad (UP) North Platte Service Area, Canadian Pacific Railway (CP) Chicago Area, New Jersey Transit (NJT), and Amtrak, who each completed a 5-year demonstration pilot at a specific point in time. More details are described in Section 6.2.1 Limitations of the Evaluation. After analyzing the data, the Evaluation Team also listed Lessons Learned for maximizing the impact and sustainability of C^{3}RS in Section 6.3.1.

The demonstration pilot sites requested that their information be kept confidential. Therefore, the information presented in this report is constrained in a number of ways:

- Names of sites are disguised when discussing data.

- Time periods of data collection are disguised because the timing of the C^{3}RS demonstration at each site was publicly known.

1.5 Organization of the Report

The report has the following sections:

- Section 2 describes C^{3}RS demonstration project.
- Section 3 covers the design, plan, and execution of this evaluation of C^{3}RS.
- Section 4 details the data collection and analysis methods.
- Section 5 provides the results of the data analysis.
- Section 6 interprets the results to formulate findings and lessons learned for railroads, and discusses the limitations of the evaluation.
- Section 7 provides the overall conclusion of the evaluation.

2. The Demonstration Project

2.1 History of the C³RS Demonstration Project

The history of the C³RS Demonstration Project is as follows:

- In 2002, FRA's RD&T HF Division decided to fund and conduct a workshop on close call reporting systems. RD&T HF requested that the Volpe Center set it up on their behalf through a government interagency agreement (Saks, 2004). One purpose of the workshop was information-sharing as a mechanism to expose railroad stakeholders to methods other industries have used for close call reporting. A second purpose was to provide a setting in which a coalition of interests could form to advance close call reporting in the railroad industry.
- To develop the workshop, in 2002 RD&T HF and Volpe formed a Planning Committee. Their task was "to decide how to introduce the railroad industry to the value of studying close calls as a way of improving safety." (Saks, 2004). The committee included representatives from carriers, the National Transportation Safety Board (NTSB), the Bureau of Transportation Statistics (BTS), as well as labor representatives from the International Association of Sheet Metal, Air, Rail and Transportation Workers' Transportation Division (SMART TD),[8] Brotherhood of Locomotive Engineers and Trainmen (BLET), Brotherhood of Maintenance of Way Employees (BWME), Brotherhood of Railroad Signalmen (BRS), and the American Train Dispatchers Association (ATDA).
- The workshop, "Improving Railroad Safety Through Understanding Close Calls," was held on April 23 and 24, 2003. Attendees included: FRA RD&T HF, 19 carriers, five labor unions, the NTSB, representatives from the airline industry, researchers, and other government transit authorities. Sessions covered lessons learned from aviation, existing rail initiatives, and breakout discussions.[9]
- In 2003, after workshop attendees enthusiastically accepted the idea of a railroad close call reporting system, FRA RD&T HF established a C³RS demonstration project. Its reasoning was that because confidential close call reporting had worked in other settings, it was worth testing to determine whether such a program would work for railroads. If they found the new approach to be effective, close call reporting could be scaled up throughout the railroad industry. RD&T HF asked Volpe Center members who had coordinated the workshop to extend their work to include planning for and implementing the demonstration project establishing the C³RS Implementation Team to do this work, which also included DIGITALiBiz and the Hile Group.
- Also in 2003, the RD&T HF established an interagency agreement with BTS. BTS was asked to establish the Third-Party procedures and system for receiving, redacting, and

[8] At the beginning of the C³RS demonstration United Transportation Union (UTU) had not yet become part of International Association of Sheet Metal, Air, Rail and Transportation Workers (SMART) Transportation Division (SMART TD).

[9]Proceedings from the FRA-sponsored human factors workshop: Improving Railroad Safety Through Understanding Close Calls. (Website). Accessed online: June 23, 2015.

performing some initial analysis on close call reports. BTS was a member of the Planning Committee and helped establish what the Third-Party could and could not do. BTS signed the original memorandum of understanding (MOU) and three of the demonstration pilots' IMOUs.

- After the workshop, the primary task of the Planning Committee was to draft a model MOU to serve as a model for the IMOUs that each C^{3}RS site would have. The model MOU was created to establish the basic parameters of how the reporting system would work based on the consensus of the stakeholders. The key was to build trust in the model, so the industry stakeholders would consider participating. The purpose of each site's IMOU was to adapt the MOU to the needs of the stakeholders at each railroad. The IMOU laid out in greater detail how the reporting system would work. It specified details regarding boundary conditions for participation and how the program would be implemented. It specified what was in scope and what was out of scope, so that stakeholders understood how to operate the program. Negotiating this model MOU involved considerable work and good will among all stakeholders, including labor, management, and FRA, because thorny issues had to be resolved. Negotiating the model MOU took 2 years.[10] Important issues about protection from discipline were resolved. The intent was to help employees become comfortable submitting information on close calls to allow the causes for those situations to be revealed, even in some situations where an employee violated a rule and was eligible for discipline. Management wanted to understand the causes of those close calls. Some of the issues the Planning Committee resolved included:
    - How would protection from discipline work? What events would qualify?
    - What would the maximum time be between an employee's observation of a close call and when the report has to be submitted to receive protection from discipline?
    - How many crew members would be protected from discipline by a specific C^{3}RS report? Only the reporter or the whole crew?
- From 2004 to 2010, FRA and the C^{3}RS Implementation Team worked to recruit railroads for the C^{3}RS demonstration pilot sites. Because the idea of a close call program in the railroad industry was novel and challenging to existing processes and culture, it took several years to find railroads that were willing to sign up. After several years on the part of the Steering Committee, four railroads were recruited: Union Pacific Railroad (UP) North Platte Service Area in 2007, Canadian Pacific Railway (CP) Chicago Territory in 2008, New Jersey Transit (NJT) in 2009, and Amtrak in 2011. Initially, transportation crafts were the focus (BLET and SMART TD), although NJT ATDA included dispatchers in addition to their train and engine crafts. Later, Amtrak added dispatchers from the Transportation Communications Union (TCU).
- In 2005, RD&T HF requested, through a government interagency agreement, that the Volpe Center conduct a Lessons Learned Evaluation. The work began with a literature review (Morell, 2006). Based on the review and an understanding of the developing C^{3}RS demonstration project, an evaluation plan was developed. Volpe contracted with

[10] C^{3}RS Publications: Memoranda of Understanding. (website) Accessed online: June 23, 2015.

Jacobs and their subcontractor partner, Syntek Technologies Inc., to provide evaluation expertise and help conduct the evaluation.[11]

- Also in 2005, the Planning Committee transitioned into the Steering Committee and membership expanded. The purpose of the group transitioned from a planning role to a strategic oversight role. The Steering Committee's purpose was to describe how C^{3}RS would operate via a model MOU; have strategic oversight of C^{3}RS; and monitor how the system was working.[12] Steering Committee meetings were open for anyone to attend. The C^{3}RS Implementation Team communicated progress to the Steering Committee and exchanged information.
- In 2009, sponsorship and funding for the C^{3}RS Implementation Team transferred from FRA's RD&T HF to the Office of Railroad Safety's (RRS) Risk Reduction Program (RRP). RRP's mission was "ensuring the safety of the Nation's railroads by evaluating safety risks and managing those risks in order to reduce the numbers and rates of accidents, incidents, injuries, and fatalities."[13] The move to RRS facilitated the recruitment of new railroads to C^{3}RS, beyond the original demonstration pilot sites. The decision to make this change in C^{3}RS oversight came from a belief that C^{3}RS was being accepted, and thus, that a transition from research to a more operational setting was warranted.
- In 2011, when Amtrak was established as the fourth demonstration pilot site, NASA began to act as a C^{3}RS Third-Party. NASA performed many of the same procedures as BTS, adapting them based on their experience with ASRS.
- In November 2014, C^{3}RS began to transition to a new FRA office, out of the RRP and to the Human Performance Division (HPD), still within RRS. The move became official in October 2015. The purpose of this move was to facilitate the national roll out of C^{3}RS, including establishing a dedicated team to administer the program, recruit, and educate railroads interested in participating in C^{3}RS.
- While the Implementation Team moved to the HPD, RD&T HF continued to sponsor the Lessons Learned Evaluation of the C^{3}RS program. The evaluation activity remained within RD&T HF because its purpose to assess C^{3}RS scientifically and objectively was more aligned with RD&T's research mission.

2.2 Description of the Demonstration Project

The C^{3}RS demonstration project consisted of several interdependent organizations with distinct roles as described in the following sections.

[11] The contractor team's company affiliation has changed several times during the duration of their work. In 2010 Jacobs acquired the team formerly owned by TechTeam Global, New Vectors LLC, and Altarum Institute. Syntek Technologies Inc. merged with Fulcrum Corporation on January 1, 2016.

[12] C^{3}RS Steering Committee Charter. November 21, 2005.

[13] US DOT FRA Risk Reduction Program. Accessed online: June, 23 2015.

2.2.1 FRA

FRA was the sponsor and funder of C^{3}RS. As described in the history section above, the Implementation Team and the Third-Party sponsorship began in the RD&T HF and moved to RRS, first in the RRP, and then to HPD. FRA sponsorship included several responsibilities:

- Hiring and funding the C^{3}RS Third-Party to confidentially collect and create the de-identified C^{3}RS report-records.
- Providing waivers to the demonstration pilot sites.
- Providing personnel for the C^{3}RS Implementation Team, both by hiring the Volpe Center and later by providing FRA RRP and HPD personnel.
- Providing inspectors from RRS from the region where each pilot resides to serve on the PRTs.
- Funding the C^{3}RS Lessons Learned Evaluation Team. (While RRS sponsored the implementation aspects of the program, the RD&T HF continued to fund and sponsor the Evaluation Team.)

FRA's RRS contains eight regional offices covering different geographical areas of the United States.[14] The RRS is responsible for executing regulatory and inspection responsibilities and includes Federal safety inspectors. To help the PRTs, FRA RRS provided regional representatives to participate as full members of the PRT at each demonstration site. The FRA PRT members had the responsibility to attend PRT meetings, help with analysis of data, and protect the confidentiality of individuals and railroads as outlined in the IMOUs. This was a requirement that was written into the IMOUs at each site. The purpose of FRA's participation on the PRTs was to provide expertise on regulations, analyze and summarize emerging trends, and help develop recommendations for corrective actions. In their training, FRA PRT members were also encouraged to help preserve the health of C^{3}RS by providing leadership and guidance.

2.2.2 C^{3}RS Implementation Team

The C^{3}RS Implementation Team began at Volpe, and later transitioned its responsibilities to RRS's RRP, and then HPD. The Implementation Team was and is responsible for training and assisting the demonstration pilot sites in their implementation of C^{3}RS. The Implementation Team also assisted the evaluation effort by providing feedback on the C^{3}RS Evaluation Team reports and participating in interviews.

The C^{3}RS Implementation Team provided training to the PRTs on MCIA problem solving. The Implementation Team also developed an MCIA software application for the PRTs to use to record their analyses. MCIA, also known as "the five whys" contributing-cause analysis, consisted of a series of questions to help identify multiple contributing factors that triggered a close call situation.[15] Some of the contributing factors a PRT was asked to consider included:

- Operational/individual factors (e.g., what demands do crew members face?)

---

[14] FRA Organizational Chart (website) Accessed online: June 23, 2015.

[15] C3RS Multiple Cause Incident Analysis (MCIA) Job Aid. (Website). Accessed online: June 23, 2015.

- Preconditions for operator acts (e.g., was the employee fatigued? Was there a miscommunication?)
- Environmental factors (e.g., was the site clean and free of tripping hazards?)
- Organizational factors (e.g., did production pressure overshadow safety concerns?)
- Outside influences (e.g., are the relevant regulations poorly written?)

2.2.3 C³RS Third-Party

The C³RS Third-Party was an independent Federal agency that collected the C³RS reports from eligible reporters from the railroads, ensured confidentiality of the reporters, and provided carrier-approved receipts showing that the employees provided reports. Given that many of the situations that were reported might otherwise result in discipline, the Third-Party ensured that reporters' confidentiality remained safe. ASRS, the aviation close call system, also uses a Third-Party as do other close call reporting systems.

BTS was the first Third-Party for the C³RS demonstration project. BTS was a legally designated "Federal Statistical Agency" and thus was able to operate under the provisions of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002. While report-records were de-identified before they were provided to the PRT, CIPSEA protection allowed some location information to remain in each report-record sent to the PRT. The security for C³RS data included a secure data storage room at each demonstration site. In addition, agreements with each railroad added an additional security measure in which the PRT physically shipped the C³RS database back and forth to BTS on a laptop once a month to obtain new report-records, thereby avoiding email transmission.

In 2010, FRA and NASA signed an Interagency Agreement describing NASA as a Third-Party for the last demonstration pilot site. Later, after the first three sites were done with their 5-year demonstration period, FRA transitioned all sites to NASA. NASA then became the only Third-Party for any new railroads to the system. NASA had extensive experience as a Third-Party with close call reporting in the aviation industry through ASRS.[16] NASA used a process similar to ASRS but modified it to allow de-identified report-records to be directed to specific railroads. NASA achieved confidentiality by removing certain information that might identify the individual submitting the report to the rail carrier. The type of information removed was variable depending on the nature of the event and the information provided. De-identification processes were accomplished through a case-by-case review by expert analysts.

Both BTS and NASA had similar report processing operations. Once the Third-Party (BTS or NASA) received the C³RS report, they initiated a process that acknowledged receipt of the report to the reporter and conducted a brief interview to obtain additional information. BTS used scripted questions for their interviews; NASA interviews were more of an open conversation that they later categorized with their pre-set codes. After the interview, the Third-Party analyst wrote up a C³RS report-record, removing personally identifying information, and classified it into a close call category. BTS and NASA used different close call classification, and NASA used the term "anomalies" instead of "categories." BTS provided C³RS report-records monthly to the

[16] NASA's Aviation Safety Reporting System (ASRS) (website). Accessed online: June 23, 2015.

PRTs to review. NASA provided daily report-records, or when Third-Party analysis was completed, through a secure portal accessed electronically.

One difference between the BTS and NASA processes was that, because of CIPSEA, BTS was able to retain and protect data in their original form and thus did not have to purge information from its records. As a result, BTS had the ability to review the original de-identified reports and analyze changes over time, or to make comparisons among the close categories of report-records. NASA achieved confidentiality by removing personally identifying information on a case-by case basis from their database and destroying original reports. NASA did not use CIPSEA. NASA reviewed CIPSEA for its applicability to this voluntary reporting model of C^3RS with OMB and NASA General Council and believed it was inadequate for maximal sharing of safety information. Additionally, NASA had the question about whether CIPSEA could apply to voluntary reported data since CIPSEA specifies "statistical" data for protection. NASA believed that the nature of voluntarily submitted data was not statistical. Besides protecting confidentiality, NASA hoped that their method would help the PRT to consider problem solving across the railroad and develop system-level solutions. A "system-level issue" was a problem in need of corrective action in which the "spread' of the problem or its corrective action crossed significant organizational boundaries. These may have been geographical boundaries, e.g., service areas or regions, or functional boundaries, e.g., "Transportation" and "Mechanical."

2.2.4 Demonstration Pilot Sites

With support from the Steering Committee, the C^3RS Implementation Team sought to recruit two passenger and two freight railroads to ensure applicability to a broad range of the railroad industry. Criteria for inclusion included: having a sufficient number of employees to achieve the Implementation Team's initial goal of 1,200 eligible reporters per site; high interest in participating; and sufficient trust between labor and management to begin the IMOU process. The start-up of demonstration pilot sites took place between 2007 and 2011. Table 1 shows which railroads joined and when. When RD&T HF established the demonstration project, they required that each site stay in the program for 5 years. RD&T HF did this because they anticipated that 5 years would be necessary to get each pilot functioning effectively.

Table 1. C^3RS Demonstration Pilot Sites

Railroad	C^3RS Demonstration Start Date	Location	Type
Union Pacific (UP) North Platte Service Area	February 2007	Yard and Road	Freight
Canadian Pacific (CP) Chicago Area	April 2008	Primarily Road	Freight
New Jersey Transit (NJT)	October 2009	Yard and Road	Passenger
Amtrak	February 2011	Started with Primarily Yard, later added Road	Passenger

2.2.4.1 UP North Platte Service Area

Union Pacific (UP) is a freight railroad that operates in North America, covering 23 states in the western two-thirds of the United States. UP has approximately 46,500 employees and 8,400 locomotives. UP has approximately 10,000 customers in the agricultural, automotive, chemical, and coal industries. UP connects West Coast and Gulf Coast ports to eastern gateways as well as to Canadian railroads and Mexico.

UP's North Platte Service Unit located in North Platte, NE, began participating in the C^{3}RS demonstration in February 2007 (Figure 4). The C^{3}RS evaluation completed final data collection in March 2013. Approximately 2,500 employees who worked on yard and road crews were eligible to submit C^{3}RS reports. The UP IMOU included road and yard members from SMART TD and BLET unions that worked within the milepost boundaries of the service unit. RRS provided inspectors from Region 6 to participate on the PRT.

UP's Bailey Yard in the North Platte service unit in Nebraska was the largest railyard in the U.S. It covered 2,850 acres and was 8 miles long, and handled approximately 14,000 rail cars every 24 hours. Three thousand cars were sorted daily in the "hump yards." The hump yards allowed 4 cars a minute to roll into any of 114 "bowl" tracks. As cars were sorted into the tracks they formed trains and headed for locations in the East, West, and Gulf Coasts of North America, as well as the Canadian and Mexican borders.[17]

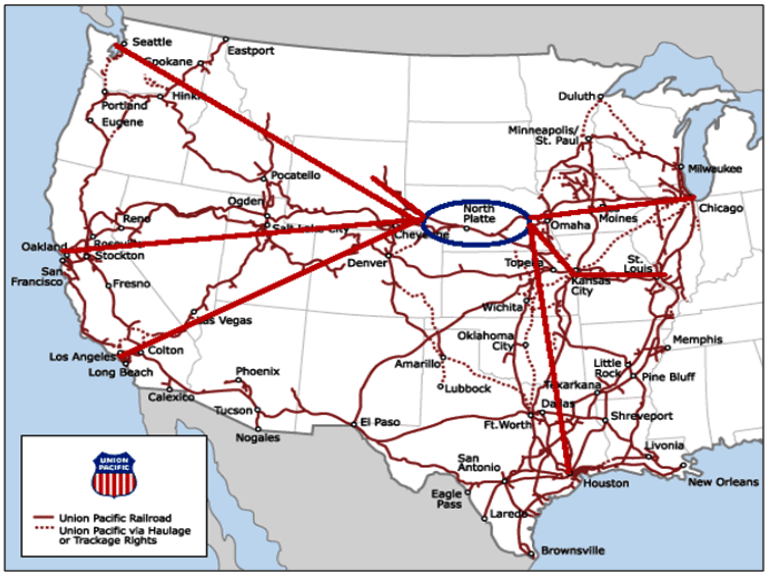


Figure 4. Union Pacific C^{3}RS Boundaries

2.2.4.2 Canadian Pacific — Chicago Service Area Road Territory

Canadian Pacific Railway (CP), a freight railroad, covers 14,000 miles from the Port of Vancouver in western Canada to the Port of Montreal in eastern Canada as well as to the U.S.

[17] Union Pacific (2014). "Bailey Yard." (Website) Omaha, Nebraska. Accessed online: August 4, 2014.

cities of Chicago; Newark; Philadelphia; Washington, DC; New York City; and Buffalo.[18] CP has approximately 16,100 employees.[19]

Members of the BLET and SMART TD unions within the CP Chicago Service Area Road Territory began participating in the C^{3}RS demonstration in April 2008 (Figure 5). The C^{3}RS evaluation completed final data collection in July 2013. Approximately 350 employees were eligible to submit C^{3}RS reports. Unlike the UP site, which contained a large yard as well as road crews, most of the CP demonstration pilot eligible reporters worked as road crews on the main line. The boundaries of the C^{3}RS demonstration pilot were defined from east to west, as the territory operated by train crews going on or off duty between Newport, MN, and Tower A-20, near Chicago, and all track in between. Crews operating in Portage and Milwaukee, WI, were covered during their entire tour of duty even while operating on track owned by other railroads. Also covered were crews working between St. Paul, MN, and Portage, WI, and crews operating to the Waterloo Pit. RRS provided inspectors from Region 4 to participate on the PRT.

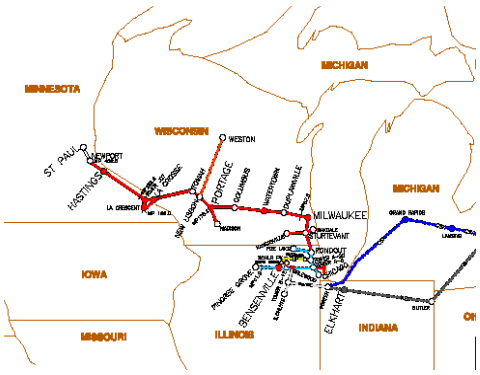


Figure 5. Canadian Pacific C^{3}RS Boundaries

2.2.4.3 New Jersey Transit

New Jersey Transit (NJT) is the State of New Jersey's public transportation corporation. NJT is the Nation's third-largest provider of bus, rail, and light-rail transit, linking major points in New Jersey, New York, and Philadelphia (Figure 6). NJT's service area is 5,325 square miles. NJT has approximately 2,027 buses, 711 trains, and 45 light-rail vehicles on 236 bus routes and 11 rail lines. NJT provides nearly 223 million passenger trips each year.[20]

NJT was the first of two passenger railroads to participate in the C^{3}RS demonstration, starting in October 2009. The C^{3}RS evaluation completed final data collection in January 2015. Approximately 1,700 members of the BLET, the American Train Dispatchers Association (ATDA), and SMART TD were eligible to submit C^{3}RS reports when NJT joined C^{3}RS. All NJT-owned and/or operated territory was part of the pilot, including the Southern Tier and Pascack Valley Line. NJT's C^{3}RS eligibility extended only to tracks owned by the company. Track used by NJT, but owned by Amtrak or Conrail, were not covered. RRS provided inspectors from Region 1 to participate on the PRT.

---

[18] Canadian Pacific (2014). "Our History." (Website) Calgary, Alberta, Canada. Accessed online: August 4, 2014.

[19] Canadian Pacific (2014). "About CP." (Website) Calgary, Alberta, Canada. Accessed online: August 4, 2014.

[20] New Jersey Transit (2014). "About Us." (Website) Newark, NJ. Accessed online: July 28, 2014.

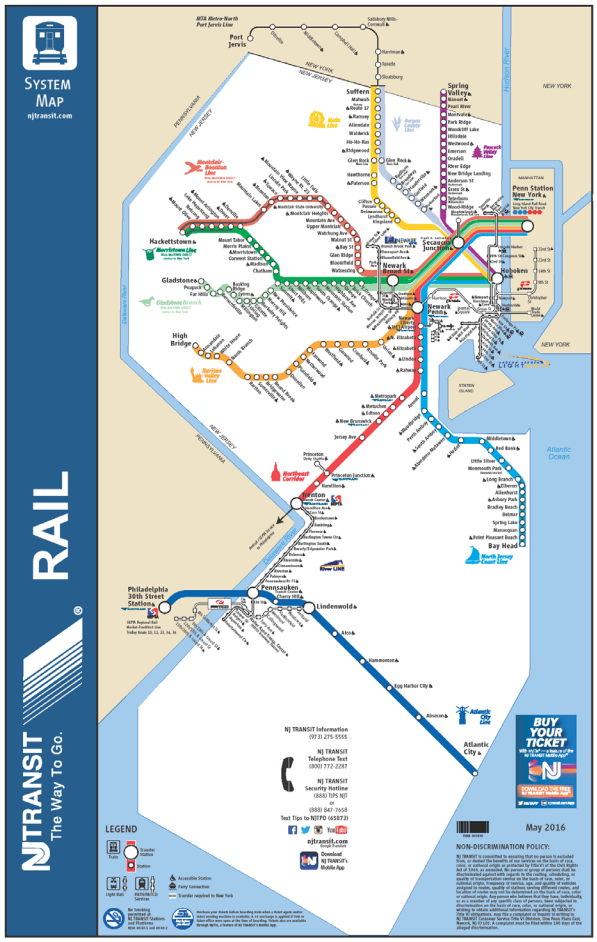


Figure 6. NJT System Map

2.2.4.4 Amtrak

Amtrak is the only nationwide passenger railroad, serving more than 500 destinations in 46 States and 3 Canadian provinces. It covers more than 21,200 miles of routes and has more than 20,000 employees. In 2011, an average of more than 831,000 passengers per weekday traveled on trains operated by Amtrak or on other railroads operating on Amtrak track.

Amtrak began participating as a C^{3}RS demonstration pilot in February 2011. The C^{3}RS evaluation completed the final data collection in December 2015. Participants included employees from BLET and SMART TD working at a majority of the yards across the United States, as defined in the IMOU (Figure 7). No road crews were included during the baseline phase, although in later years their C^{3}RS program was expanded. Approximately 1,400 employees were eligible to submit C^{3}RS reports when it began in 2011. Amtrak employees in the Northeast Corridor (PRT East) yards and in the Chicago-, Miami-, Seattle-, and Los Angeles-area (PRT West) yards could report close call events to NASA at the beginning of their demonstration. The Oakland yard was added several months later. Dispatchers from Transportation Communications Union (TCU) in Chicago, additional yards, and road employees were added in 2013. The C^{3}RS Evaluation at Amtrak focused on the West PRT. RRS provided inspectors from multiple regions to participate in the PRTs.

Figure 7. Amtrak C³RS Boundaries

2.2.4.5 Disguising the Identities of Sites in this Report

The demonstration pilot sites asked that their respective data be kept confidential; therefore, when describing results, the labels for the sites were disguised (Sites 1, 2, 3, and 4), and the assignment of labels is not the sequence in which the sites joined the research project, which is public information. However, to avoid confusion, site designations are consistent throughout the report, so Site 1 always refers to the same demonstration pilot site; Site 2 always refers to the same site, and so forth. The assignment of numbers is consistent with past evaluation reports.

2.2.5 Demonstration Pilot Site Personnel

Within each demonstration pilot site, there were three groups of personnel involved in C³RS: the eligible reporters, the PRT, and the Support Team.

2.2.5.1 C³RS-Eligible Reporters

Eligible reporters were employees that could submit a C³RS report. An employee's eligibility was determined by the scope of the IMOU at each railroad that was negotiated between labor, management, and FRA. Reporters were responsible to recognize close call events, submit reports to the C³RS Third-Party, and participate in interviews with the Third-Party. In all four sites, the transportation crafts within the BLET and SMART TD unions were covered. Over time, additional crafts were added such as dispatchers, American Train Dispatchers Association (ATDA) at NJT, and Transportation Communications Union (TCU) at Amtrak.

Some eligible reporters participated in voluntary C³RS Evaluation interviews as well.

2.2.5.2 Peer Review Team

The PRT was responsible to analyze C³RS-report-records provided by the Third-Party to identify why the unsafe event occurred and provide recommendations for corrective actions. The PRT used the MCIA method to create C³RS-analyzed-cases, as they were trained by the Implementation Team. (For more information on the MCIA method, see Section 2.2.2.) After the PRT determined the causes of a given C³RS close call event through the MCIA problem-solving

process, they developed recommendations for corrective actions for their railroad. The PRT then sent recommendations for corrective actions to appropriate managers or the designated Support Team.

There were several tools used by the PRT during the demonstration pilots. First, there were secure laptops. For three of the demonstration sites, BTS loaded C^{3}RS-report-records onto a secure laptop and physically mailed it to the PRT once a month. Each report-record included de-identified information about a close call event, including some classification information created by BTS. Classification items included the close call category and whether or not the reporter was eligible for discipline protection. Once a month, the PRT mailed the laptop back to BTS to get more de-identified report-records. Secondly, the PRTs used an FRA-provided MCIA software application to analyze and record their work. Even though Amtrak received their report-records from NASA through a secure electronic transfer, they received a copy of the MCIA software application to use in their analysis. Finally, the PRTs also developed their own spreadsheets and documents to record their analysis.

The PRT assisted the Evaluation Team by helping set up interviews and participating in feedback sessions. They also served as respondents in the interview data collection.

2.2.5.3 The Support Team

Three of the four sites had a Support Team, comprised of managers at the demonstration site and/or senior managers. This team was responsible for reviewing the recommendations from PRT and implementing changes to improve safety. In one of the demonstration sites, one or two managers had this responsibility. At the other three sites, a formal cross-function Support Team was established. The original C^{3}RS design did not call for a formal Support Team. As C^{3}RS rolled out, the Implementation Team recognized the need for such a group.

The Support Team assisted the Evaluation Team by approving and providing points of contact for the evaluation data collection. In addition, they served as respondents in the interview data collection.

2.2.6 Lessons Learned Evaluation Team

RD&T HF sponsored an independent, systematic evaluation of C^{3}RS. It was important to RD&T that the C^{3}RS Lessons Learned Evaluation Team be independent of the C^{3}RS Implementation Team, i.e., conducted by people who were not involved in implementing the program, to provide an unbiased, scientific assessment of C^{3}RS. To conduct the evaluation, RD&T established an interagency agreement with the Volpe Center. The Volpe Center hired Jacobs Engineering Group (Jacobs) and their subcontractor partner, Syntek Technologies Inc., to assist with the evaluation. In addition, the Bureau of Transportation Statistics (BTS) was contracted to collect survey data for evaluation for three of the four sites. The Evaluation Team was responsible for conducting a scientifically credible evaluation, including: data collection; analysis; interpretation; stakeholder engagement; and distribution of findings to each site and eventually to the industry.

2.3 How the C^{3}RS Works Across the Industry and Within the Demonstrations

Figure 8 shows the "logic model" that identifies C^{3}RS's stakeholders; key activities; first, second, and third order impacts; and shows the relationships among them. Logic models are

pictorial representations that identify relationships between what a program does, and its long- and short-term goals. They are also valuable tools to identify what needs to be measured and relationships among measures. They are helpful in formulating an overall evaluation methodology. Finally, logic models are used to develop consensus and a common understanding of what data need to be generated by the evaluation (Rogers, 2005). This logic model depicts the relationships among stakeholders, C^{3}RS program activities, and their consequences. Figure 8 includes both rows and columns. There is a row for each type of C^{3}RS stakeholder. The first two columns on the left show activities that occur during the initial implementation of C^{3}RS. The next columns show the first, second, and third order impacts from left to right.

In Section 2.2, the roles and responsibilities of the stakeholders were described. The logic model provides an explanation of what interim outcomes and eventual outcomes are expected to be observed as a result of the stakeholder activities in C^{3}RS. In the logic model, there is a row for each stakeholder group:

- The FRA row refers to all the involved offices in FRA and its C^{3}RS Implementation Team.
- The Third-Party row refers to both BTS and NASA.
- The Demonstration Pilot Site area contains four rows: Support Team, PRT, C^{3}RS-Eligible Employees, and Multiple Stakeholders (which refers to a combination of the three previous groups).
- The Railroad Industry row refers to both carriers and national labor (e.g., BLET, SMART TD, ATDA, and TCU).

In Column 3, the potential first order outcomes are as follows:

- As employees use the system and there are no disciplinary repercussions, trust builds and the system is used more.
- As more people submit reports, the Third-Party (BTS or NASA) processes more reports. Therefore, BTS/NASA is required to plan for more resources to accommodate the increased reporting.
- In turn, the PRT starts to analyze data from more report-records.
- The Support Team implements corrective actions intended to improve safety, possibly in collaboration with other safety programs. These improvements are communicated to the eligible reporters and the carrier.
- Employees involved in C^{3}RS improve their safety practices and safety awareness.

In Column 4, the potential second order outcomes include:

- The Third-Party provides reports of trends which FRA can analyze further on its own. To be more efficient, the Third-Party starts to sample reports for follow-up interviews, instead of doing a full interview for every report.
- The Support Team uses the C^{3}RS data to make more improvements across the company.
- Safety and safety culture improve (which will later be referred to as an “impact” to be measured in the evaluation).

In Column 5, the potential third order outcomes are:

- FRA's greater understanding of safety risks impacts policies and decision making.
- Eligible reporters' awareness of improved communication and safety increases their morale.
- Improvements in safety lead to decreases in costs and improvements in productivity.
- The railroad industry as a whole has knowledge of best safety practices and improves safety culture.

One impact that may occur earlier, during implementation, is related to the development of the IMOU at each C^{3}RS site (Column 1). During the IMOU process, C^{3}RS Implementation Team, on behalf of the FRA, works with the local FRA inspectors, participating union representatives (local and national), and railroad management to negotiate an agreement that sets out the rights and obligations for the stakeholders involved. The process of negotiating the IMOU with the various parties builds trust and willingness to use a new way of identifying and resolving safety issues. As a result of the IMOU negotiation, the safety culture may change at this early step (Lewin, 1947). Lewin coined the terminology *unfreezing – change – refreezing* for safety culture to depict the fluid nature of the progress. This initial shift in the safety culture was also seen in FRA's study on participative rules revision, where management became active and visible in its commitment to building a participative culture before the rules revision process began (Ranney, 2004).

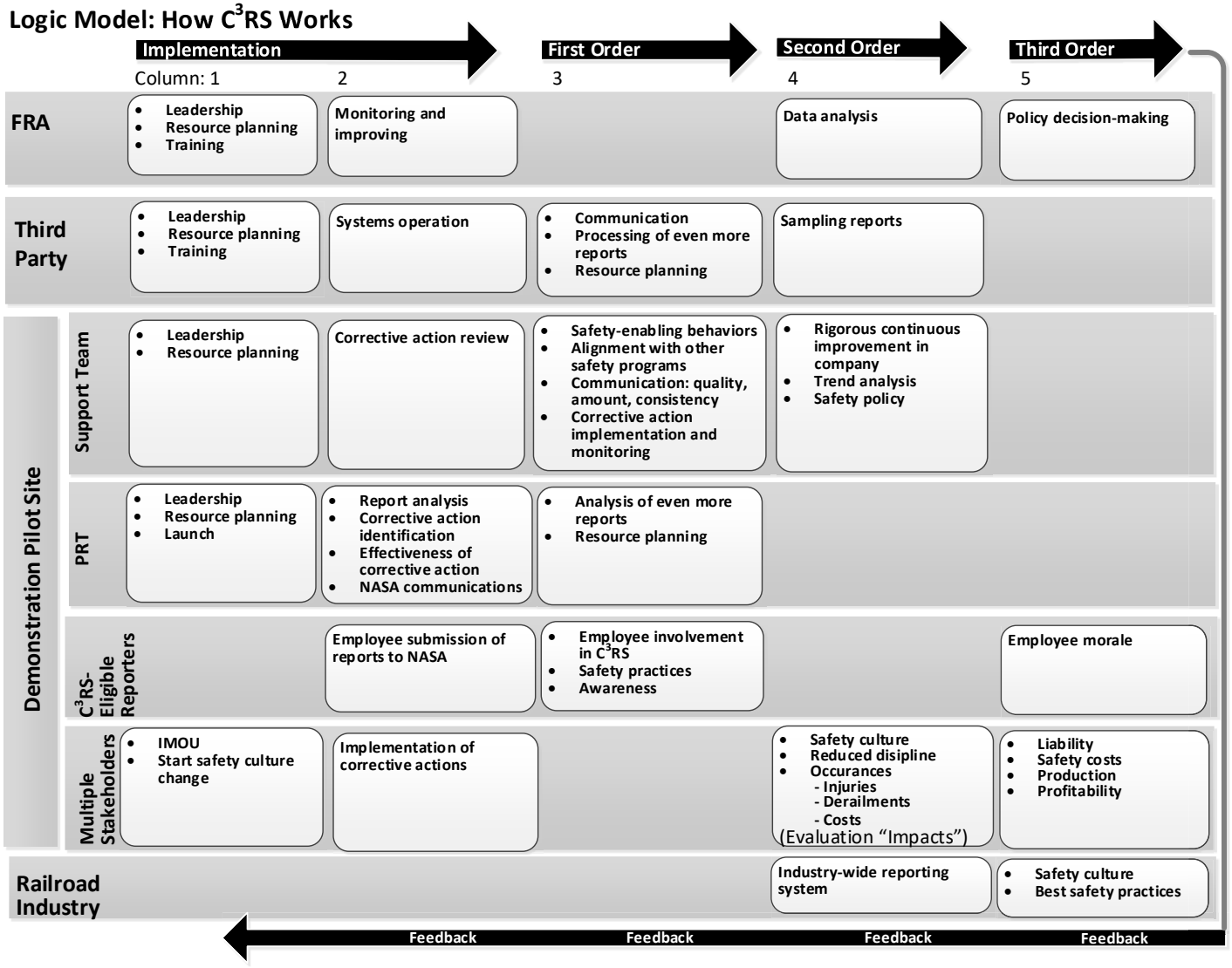


Figure 8. C³RS Stakeholder Logic Model for Industry and Demonstration

3. Evaluation

The C^{3}RS Lessons Learned Evaluation is both formative and summative, and it assesses the four C^{3}RS demonstration pilots. The evaluation includes multiple qualitative and quantitative measures in a mixed method design to assess the implementation, impact, and sustainability of the demonstration pilots. Each demonstration was studied for approximately 5 years.

3.1 Evaluation Questions

To assess C^{3}RS's potential for the railroad industry; the Lessons Learned Evaluation was designed to answer three major questions.

Table 2. Evaluation Questions

Evaluation Type	Evaluation Questions
Implementation	What conditions are necessary to implement C^{3}RS as planned in a demonstration?
Impact	What is the impact of C^{3}RS on safety and safety culture?
Sustainability	What factors help to sustain C^{3}RS long-term, beyond the demonstration?

The implementation question addresses the conditions that led to C^{3}RS's implementation at the demonstration sites as planned. In addition, if complete implementation is achieved, then it becomes justifiable to assess the impact of the program (Patton, 1997).

The impact question addresses how C^{3}RS affects its intended goals (Rossi, Freeman and Lipsey, 1999). In this case, the intended impacts are improved safety and safety culture.

The sustainability question addresses what factors led to the demonstration pilots being sustained after the conclusion of their demonstration periods. It also addresses the long-term sustainability of C^{3}RS in the industry and the role of FRA.

The inclusion of a focus on implementation is important as implementation failure can often be confused with failure to show impacts. If a program was not implemented adequately, then there may be no observed bottom-line benefit. RD&T HF wanted to know if any failure in impacts observed came from the design of the program or the way it was implemented. If no impacts were observed after a sufficient amount of time, it could be for (at a minimum) two different reasons. First, failure could be caused by the way a given site implements C^{3}RS and not the program itself per se. Second, C^{3}RS could fail if it is not viable for the railroad industry. Assessing the implementation at four demonstrations provides sufficient information on the first reason. Assessing the degree to which the first reason occurred makes it easier to determine if the second reason occurred.

3.2 Evaluation Design

3.2.1 Formative and Summative Evaluation

The evaluation of the C^{3}RS demonstration pilots is both formative and summative. The purpose of the summative component is to provide scientifically credible conclusions about the ability of C^{3}RS to be successful in the railroad industry setting. This final report includes summative conclusions. The purpose of formative evaluation is to assess if the demonstration pilots are being implemented as planned and provide recommendations for improvement.

To maximize both the likelihood of good implementation, as well as the overall usefulness of this evaluation (Patton, 1997), the evaluation includes a formative component (Rossi et al., 1999) in which the Evaluation Team assesses whether the demonstration pilot is being implemented as planned. The Evaluation Team periodically and directly informs site stakeholders of its assessment of the demonstration pilot and provides recommendations for improvement. More information on the feedback sessions is included in Section 3.4.2. The formative evaluation provides information about the realities of implementation, beyond what was expected when the implementation was planned.

3.2.2 Mixed Method Evaluation

When an experimental design with random assignment is not feasible, the most rigorous research design uses both quantitative and qualitative data collection and analysis methods from multiple data sources (GAO, 2009). In social science research, the concept of combining qualitative and quantitative data into a research design is called "mixed methods." One common definition of mixed methods is provided by Creswell (2014):

> *…(mixed methods research is) …an approach to research in the social, behavioral, and health sciences in which the investigator gathers both quantitative (close-ended) and qualitative (open-ended) data, integrates the two, and then draws interpretation based on the combined strengths of both sets of data to understand research problems. A core assumption of this approach is that an investigator combines statistical trends (quantitative data) with stories and personal experiences (qualitative data). This collective strength provides a better understanding of the research problem than either form of data alone.*

Three methods from Creswell (2014) in use in the C^{3}RS evaluation are described in Table 3: merging, building, and explaining.

Table 3. C^{3}RS Evaluation Mixed Methods

Creswell Mixed Method Types	How the C^{3}RS Evaluation Uses this Method
Merging – compare the qualitative and quantitative data to each other	Determine implementation strength from both interviews (qualitative) and the frequency of C^{3}RS-analyzed-cases in each close call category (quantitative).

Creswell Mixed Method Types	How the C³RS Evaluation Uses this Method
Building – use qualitative data to build justification for quantitative analysis	If interviews and corrective action data suggest corrective actions are being implemented (qualitative) then analyze safety data (quantitative) to see if corrective actions influence impacts.
Explaining – use qualitative data to explain the results of the quantitative data	If safety data show both improvements and degradation over time (quantitative), then use reporting and interview data (qualitative) to build a timeline to assess if implementation strength coincides with the safety impact trends.

Qualitative Emphasis

As shown in the first row of Table 3, different types of qualitative and quantitative data can be checked for merging to see if measures of program activity support one another. Qualitative data are also used to both build a justification for the quantitative analysis and explain changes over time (the second and third rows of Table 3).

The design for the analysis of qualitative data uses a case-study methodology, which is described below (Yin, 2003).

Quantitative Emphasis

Quantitative analysis focuses on impact, and thus is not employed until qualitative analysis indicates that impact is a reasonable expectation ("Building" in the second row of Table 3.) The quantitative data are analyzed in a quasi-experimental design (Shadish, Cook, and Campbell, 2002), in which performance is measured and evaluated following the application of C³RS to the demonstration site. For safety-related occurrence data, available demonstration pilot performance is compared to sites without C³RS to see if something unique is happening in the sites with the treatment.

Case Study Methodology

One way to characterize the methodology described here is as a multiple comparative case study, with each case consisting of rich, in-depth analysis of 5 years' worth of extensive qualitative and quantitative data (Yin, 2014). In the context of the case study methodology, each demonstration pilot site is seen as a case. The analytical challenge is to determine what it is about each case (or each site) that makes its outcome similar or different from the others.

Cases were chosen by means of a "structured convenience sample." The choices were "structured" with respect to the research methodology (two freight railroads and two passenger railroads). Within this methodology, the sample was considered a "convenience" in the sense that the C³RS implementation team spent several years making concerted efforts to recruit the participating railroads, succeeding with four and failing with three others. Therefore, the sample is made up of carriers that chose to join the C³RS demonstration project. The sample assured coverage of a wide range of railroad functioning. The necessity of relying on volunteer railroads, however, means that caution is needed in extending conclusions to smaller passenger railroads or non-Class 1 railroads.

3.3 Evaluation Plan

To answer the three evaluation questions, the Evaluation Team developed one model with three subsections that depicts the theory of action by which C^{3}RS translates its implementation activities into desired impacts. Figure 9 identifies the activities, time sequences, actors, impacts, and relationships that need to be assessed in order to address the three evaluation questions.

The Theory of Action Model in Figure 9 is divided into broad sections: initial implementation (green/medium gray); advanced implementation (blue/dark gray), and impacts (yellow/light gray). Initial implementation activities include: eligible railroad employees submitting reports to BTS or NASA, the reports being consolidated and provided to the PRT, and PRT problem solving. Advanced implementation activities would be observed if the initial activities were executed sufficiently and include implementing corrective actions concerning relatively minor issues, followed by corrective actions to address more serious issues. If the initial and the advanced activities were implemented as planned, then the impacts to safety and safety culture would be observed.

In addition to the three sections, the rectangles at the bottom of the Theory of Action Model show elements that continue throughout the entire C^{3}RS implementation process. They represent factors outside the program that may affect the program's sustainability. These include elements such as the social, economic, and policy environments that may affect the program. Feedback arrows are not included for these rectangles because they are thought to have an ongoing, diffuse, and potentially powerful influence on the rest of the model.

Numbers in each element of the model constitute an index for the Evaluation Team to help sort variables and organize the analysis. Due to the complications of real-world data collection, not every element of the model is measured equally or at all. However, even when an element is not measured, its existence serves as a guide to where data are strong and weak with respect to testing the model.

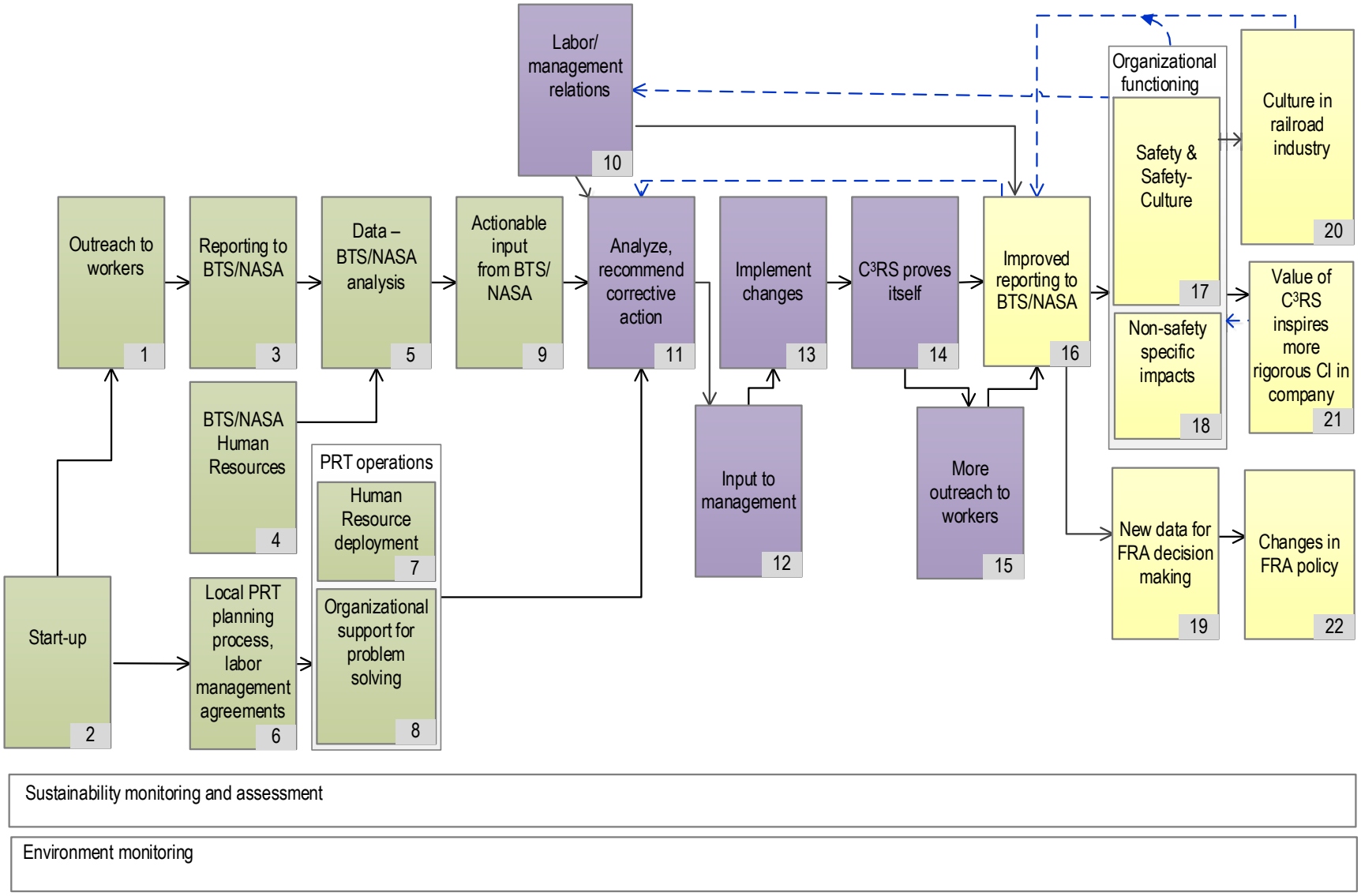


Figure 9. Theory of Action Logic Model

3.3.1 Implementation Evaluation

The goal of the Implementation Evaluation is to answer the question, "What conditions are necessary to implement C^3RS as planned in the railroad industry?" As used in this evaluation, the Implementation Evaluation serves multiple purposes. It provides:

- Rapid feedback to program implementers to help them determine midcourse corrections.
- Guidance to people contemplating similar programs in the future.
- Justification for performing impact evaluation.

To supply this feedback and guidance, the Implementation Evaluation also asks such sub-questions as:

- Is the program being implemented according to plan?
- Are deviations from the plan articulated clearly and implemented effectively and usefully?

In Figure 10, the area of the Theory of Action Model related to Implementation Evaluation is enclosed in the red/dashed box. The advanced implementation activities (blue/dark gray) would be observed if the initial activities (green/medium gray) are performed as planned. For example, C^3RS reporting to BTS /NASA (Box 3) demonstrates that employees are submitting reports. The completion of MCIA analysis by the PRT and creation of corrective action recommendations are the first indicators of advanced implementation (Box 11). Evidence that corrective actions have been reviewed and implemented also demonstrates advanced C^3RS project implementation (Boxes 12 and 13) as planned. Indicators of other implementation activities may also be measured. After it is determined that a site has completed the basic implementation activities in the Theory of Action model, implementation strength can be evaluated and compared to other sites.

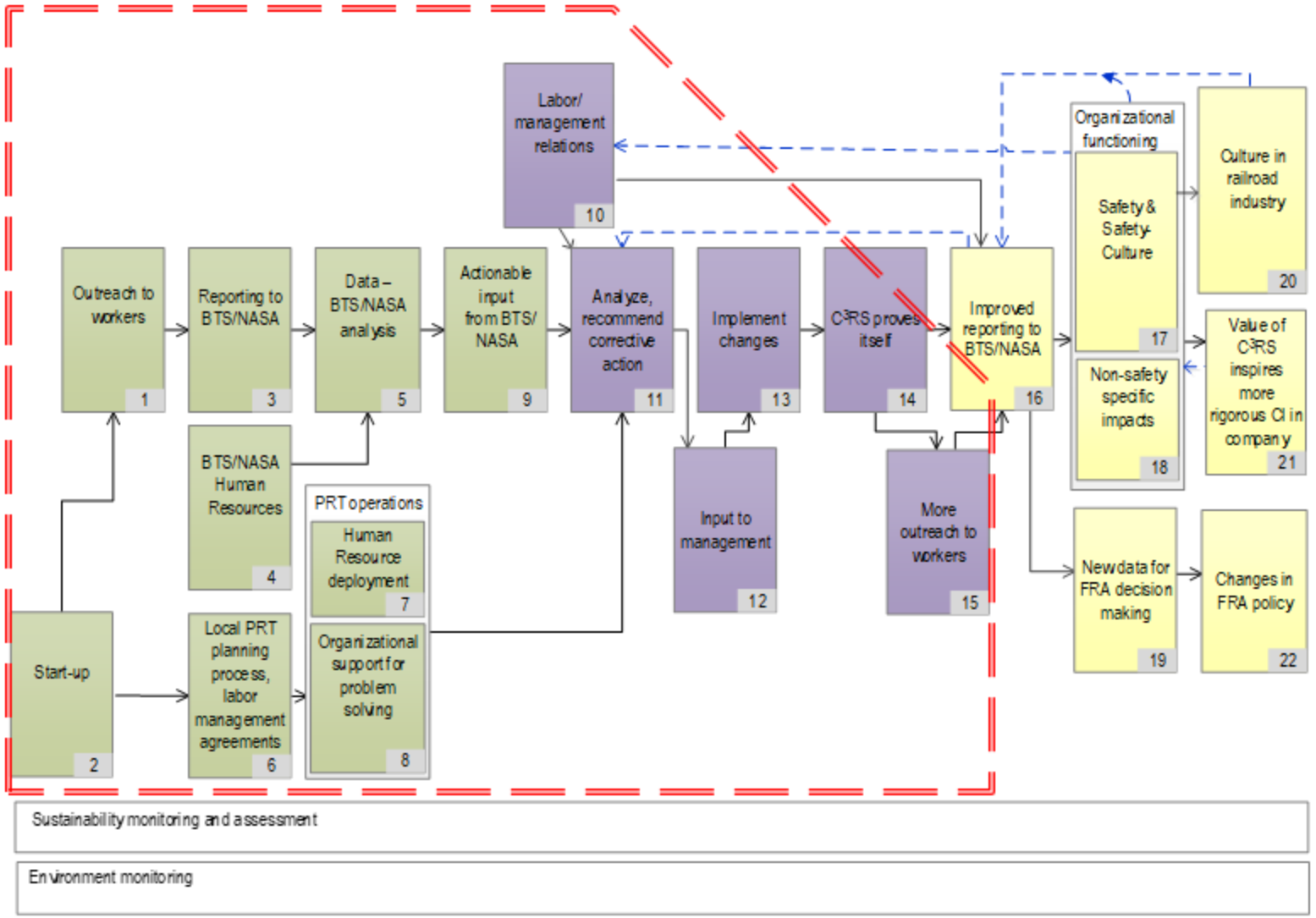


Figure 10. Theory of Action Logic Model: Implementation Activities (Initial and Advanced Activities)

3.3.2 Impact Evaluation

The goal of the Impact Evaluation is to answer the question: "What are the impacts of C^3RS on safety and safety culture?" For this evaluation, Impact Evaluation focuses on both intended and unintended short-, intermediate-, and long-term impacts. Of course, "unintended" impacts cannot be specified ahead of time and so are not in the model. However, the evaluation plan used to test the model also includes checking for unintended consequences.

Impacts cannot be assessed if the program implementation is weak. It is a well-known pitfall in the field of program evaluation that implementation failure is often confused with program failure. To avoid this risk, the Implementation Evaluation is conducted and if the demonstration site's implementation is not as planned and not sufficient then an impact evaluation is often not warranted.

The Impact Evaluation question can be decomposed into three sub-questions:

- What has happened as a result of the program having been implemented?
- What were the short-, intermediate-, and long-term impacts?
- Were there any unexpected or unintended impacts?

The impact activities region (yellow/light gray) in the Theory of Action model identify all the impacts that C^3RS might reasonably be expected to influence (Figure 11). However, time and resource constraints on the evaluation prevented analysis of all possible impacts. The evaluation focused on collecting impacts at the demonstration sites, which is marked in the red/dashed box in Figure 11 and includes boxes 17 and 18.

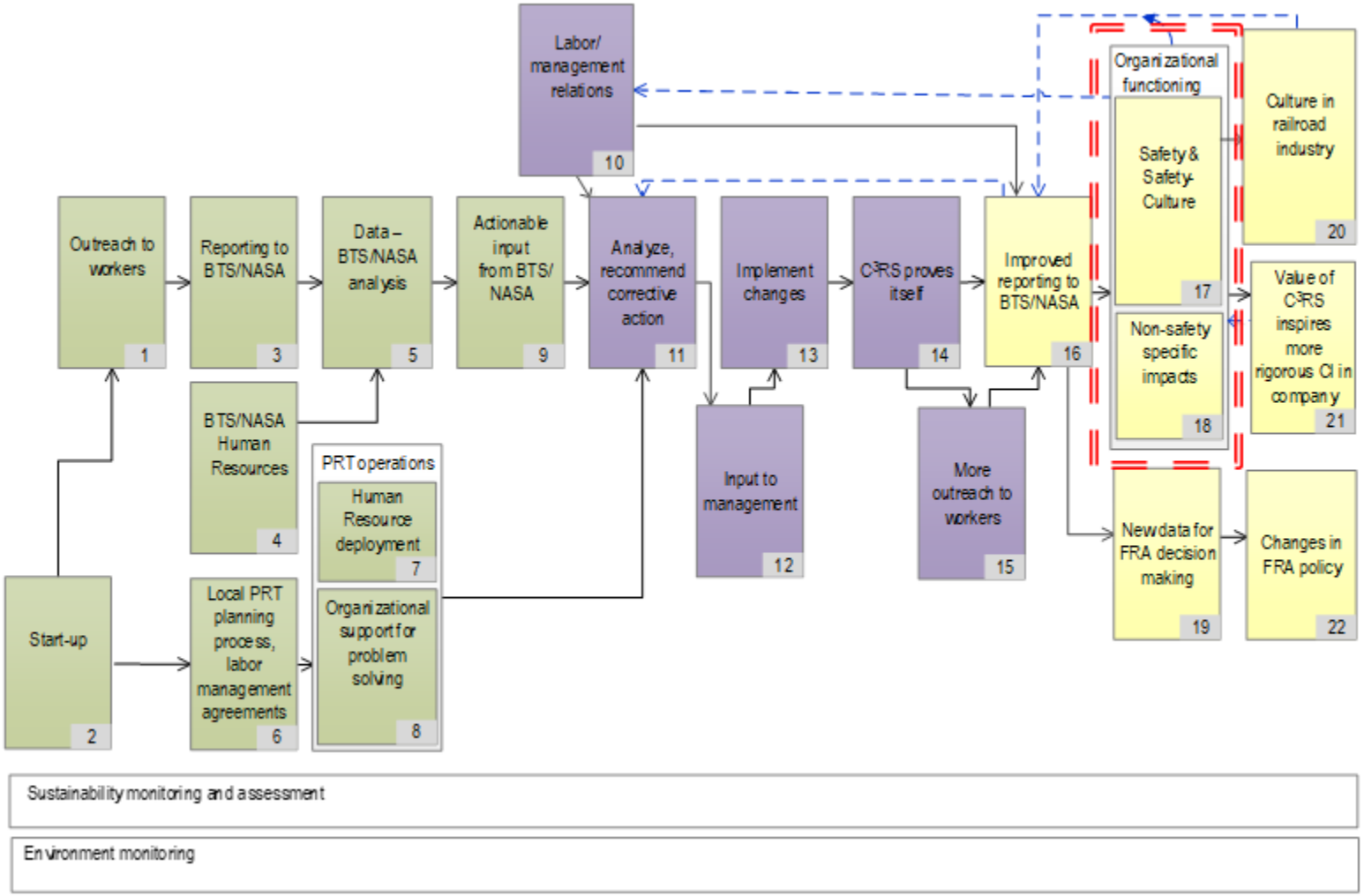


Figure 11. Theory of Action Logic Model: Impacts Measured

3.3.3 Sustainability Evaluation

The goal of the Sustainability Evaluation is to answer the question, "What factors help to sustain C^3RS long-term, beyond the demonstration?" Sustainability for a demonstration pilot site means the continuation of C^3RS at that site, and any expansion of its reach within that site or carrier. Sustainability Evaluation also covers the expansion and long-term usage of C^3RS in the railroad industry beyond the demonstration. Examining the evidence of what contributes to sustainability is needed because successful implementation and successful impact do not guarantee longevity. Specific evaluation activities are needed to answer questions such as:

- What contributed to C^3RS expanding beyond the initial scope of four demonstration sites? What contributed to C^3RS expanding to other divisions, geographical areas, or crafts within the carrier?
- What contributed to C^3RS continuing at the demonstration sites after the initial evaluation period of 5 years?
- What contributed to FRA expanding C^3RS to other railroads beyond the demonstration sites?
- How can continuation and/or extinction observed be explained?

The rectangle labeled "Sustainability monitoring and assessment" on the bottom of the Theory of Action model indicates that issues affecting sustainability can arise at any point in the C^3RS program, as indicated by the red/dashed box in Figure 12.

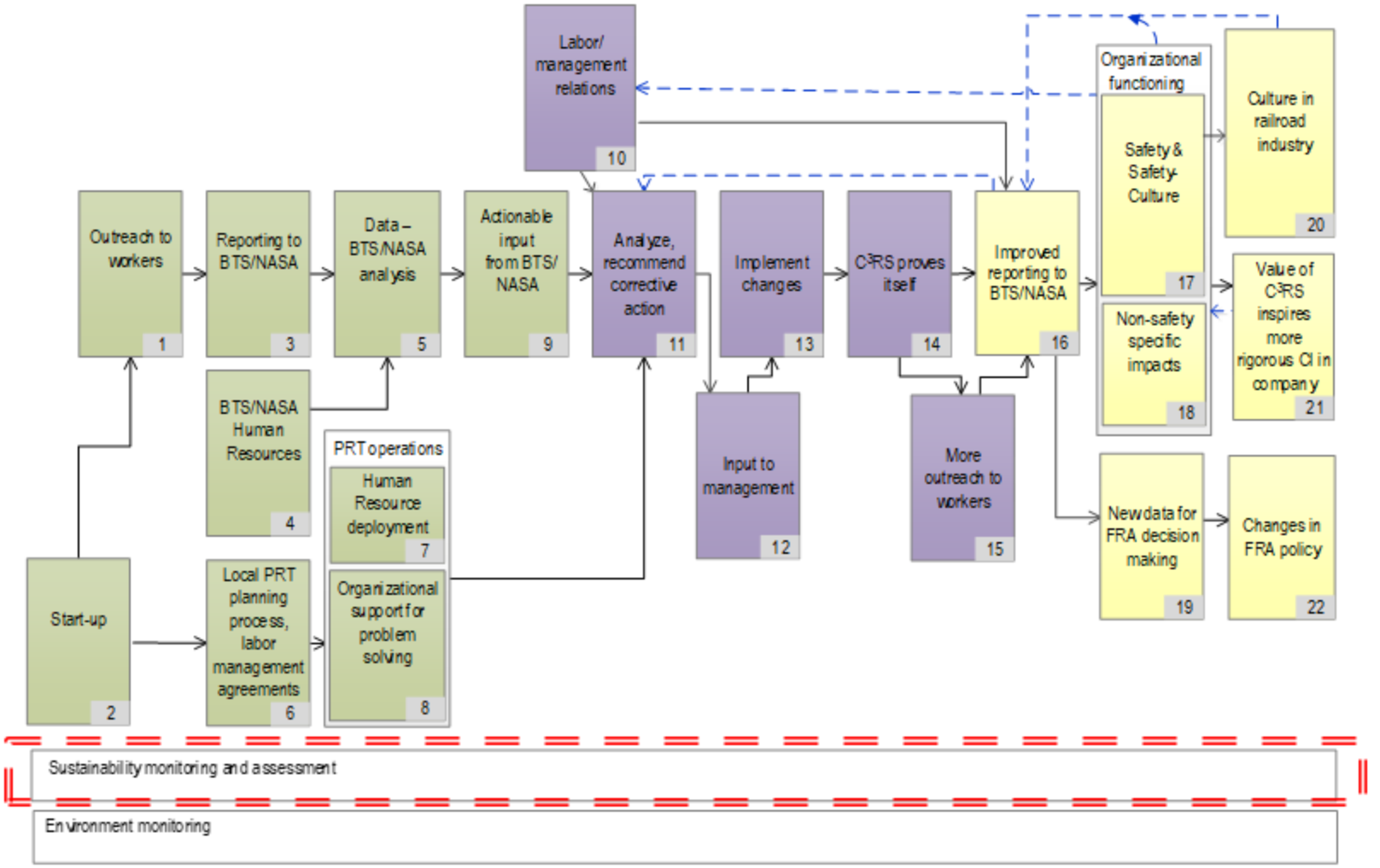


Figure 12. Theory of Action Logic Model: Measurable Elements in Sustainability Evaluation

3.4 Industry and Demonstration Pilot Stakeholders and Methods of Engagement

3.4.1 Evaluation Stakeholders

The C^{3}RS Evaluation provides information to the stakeholders within the demonstration pilot sites and across the industry. Below is an explanation of how each stakeholder benefits from the evaluation:

- **FRA:** The mission of RD&T is to conduct research and development efforts that improve safety in the railroad industry. As such, it needs scientifically credible information to provide information to RRS and the industry. The Impact Evaluation offers insights on the impacts at the demonstration sites to assess benefits and to justify budget requests to Congress. RRS needs the results from the evaluation in order to make decisions about moving C^{3}RS from a research program to an ongoing safety program.
- **C^{3}RS Implementation Team:** Implementation Team members from the Volpe Center and RRS HPD need feedback about the effectiveness of their efforts to ensure success at the four C^{3}RS demonstration pilot sites as well as any sites that may join C^{3}RS in the future. The Implementation and Sustainability Evaluation helps the Implementation Team determine how to implement the C^{3}RS program within the industry in a way that is successful and sustainable.
- **C^{3}RS Third-Party—BTS and NASA:** The Third-Party agencies need feedback about the effectiveness of their efforts to ensure success at the four C^{3}RS demonstration pilot sites and in any future C^{3}RS sites. With feedback, they can make mid-course corrections.
- **Demonstration Pilot Support Team:** Managers on the Support Team and other senior managers from the demonstration sites need answers to the evaluation questions to aid in their internal decision-making about the program. When they agreed to participate in the demonstration project they were reassured that they would receive feedback on the progress at their site in exchange for helping to fine-tune the safety program. Their decision making includes whether to support the program's expansion more broadly across their railroad on the basis of the corrective actions effectiveness and efficiency.
- **Demonstration Pilot PRT:** The PRT needs feedback on the effectiveness of its activities. It needs the evaluation process as a mechanism through which it can confidentially hear what the rest of the workforce is suggesting in terms of PRT improvements, as well as share their own perspectives about the effectiveness of other C^{3}RS stakeholders, such as the C^{3}RS Implementation Team, the Third-Party, and railroad management. The labor officials serving on the PRT need the evaluation to demine how to promote C^{3}RS among their membership. The labor unions included in the C^{3}RS demonstration were: BLET, SMART TD, ATDA, and TCU.
- **Demonstration Pilot C^{3}RS-Eligible Reporters:** Front-line employees who are eligible to report need a method in which they can confidentially share their views about how C^{3}RS is working at their site, including its strengths and weaknesses. The evaluation with its data protection provisions supplies such a method.
- **Railroad Industry – Labor Organizations and Carrier Management:** Labor and management throughout the industry need the evaluation results to help them make

decisions about whether to support C^{3}RS. The C^{3}RS Steering Committee was a group of railroad industry stakeholders in the earlier years of C^{3}RS. In later years, FRA RD&T began supporting a C^{3}RS User Group and providing updates to the Transportation Review Board (TRB) Rail Operational Safety Committee (AR070), in place of the Steering Committee. The C^{3}RS User Group was formed in 2012 to allow the PRTs across the demonstration sites to share non-confidential information. Yearly User Group meetings included management and labor representatives from the demonstration sites, BTS, NASA, the Implementation Team, and FRA RD&T HF and RRS. The TRB Rail Operational Safety Committee includes members from carriers; national labor, such as BLET and SMART TD; educational organizations; and railroad associations.

3.4.2 Stakeholder Engagement

The C^{3}RS Evaluation interacted with its stakeholders to present and refine evaluation findings; obtain insights as to how the Evaluation Team should proceed; and improve implementation, impacts, and sustainability of the program. Four engagement methods were used.

- **FRA Reviews:** Stakeholders at FRA were engaged in the evaluation in multiple ways. The Evaluation Team provided the program manager at the FRA RD&T HF with monthly progress reports and regular meetings concerning the evaluation status. FRA RD&T HF reviewed all publications, briefings, and presentations. In addition, senior managers at FRA RD&T and RRS received specially focused presentations.
- **C^{3}RS Implementation Team Reviews:** A sub-set of the C^{3}RS Implementation Team, initially from the Volpe Center and later from RRS HPD, completed reviews of drafts of all evaluation briefings and reports. These reviews were valuable to the evaluation because of the Implementation Team's in-depth knowledge of the way C^{3}RS was being implemented. In particular, their knowledge contributed to decisions about when to collect additional data, interpretations of the information, and how to present some findings. With respect to evaluation use, their interaction with the Evaluation Team alerted the Implementation Team to issues that influenced the implementation of C^{3}RS.
- **Phased Briefings to Demonstration Site PRTs, Support Teams, and Eligible Reporters:** The Evaluation Team provided each demonstration pilot site's PRT and Support Team with briefing presentations of results unique to their railroad at baseline, midterm, and final evaluation phases. These briefings were often conducted in person and included evaluation methods, findings, and recommendations. Audiences included PRTs and Support Teams. These briefings benefited both parties. The Evaluation Team received critique and useful additional information. The railroads received information that helped them improve their C^{3}RS activities.

  PRTs and Support Teams provided feedback to eligible reporters based on the evaluation briefings.
- **Communication to Railroad Industry:** The Evaluation Team provided communication to the Railroad Industry through: the C^{3}RS User Group, publicly available reports, and public conference presentations. At the User Group, because of the multi-railroad audience, presentations had to eliminate company confidential, railroad-specific information, and concentrate on findings that are more generally useful. The Evaluation Team also wrote Research Results for the FRA website, which were reviewed and

published by the FRA RD&T (Federal Railroad Administration, 2008, 2012, 2013, 2014, 2014, 2015, 2016, and 2016). Research Results were written about each site after each phase of evaluation and also contain some discussions of results across the sites. Conference presentations were made to the TRB Rail Operational Safety Committee (2012–2017) and the American Evaluation Association (2013–2016). Prior to being made public, all public reports were reviewed by senior management at each demonstration pilot site whose data are included; labor and management members of all the PRTs; the third parties (BTS and NASA) if their data are used; the FRA RD&T HF; and the C^{3}RS Implementation Team.

3.5 Program Evaluation Standards

In 2013, the RD&T established a foundation for guiding systematic, improvement-oriented evaluations and institutionalizing program evaluation.[21] While the C^{3}RS Evaluation began in 2006, before the Evaluation Implementation Plan was created, it does comply with the standards as described in the C^{3}RS Evaluation Standards Attestation Form (Appendix A. Evaluation Standards Attestation Form). The summaries of ANSI-approved standards were drawn from the Joint Committee on Standards for Educational Evaluation,[22] and reprinted with the Committee's authorization and adopted for use by FRA's RD&T. The standards include:

- **Utility**. Utility standards are intended to ensure that an evaluation will serve the information needs of the intended users.
- **Feasibility**. Feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.
- **Propriety**. Propriety standards are intended to ensure evaluations will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.
- **Accuracy**. Accuracy standards are intended to ensure that an evaluation will reveal and convey valid and reliable information about all important features of the subject program.
- **Evaluation Accountability**. Evaluation accountability standards are intended to ensure that those responsible for conducting the evaluation document and make available for inspection all aspects of the evaluation that are needed for independent assessments of its utility, feasibility, propriety, accuracy, and accountability.

3.6 Internal and External Reviews

The evaluation analyses were subjected to internal and external review. The primary focus of the internal and external reviews was to verify the scientific credibility of the statistical methods used to analyze occurrence data.

---

[21] "Evaluation Implementation Plan: Office of Research & Development" FRA/ OSD (Nov 2014). DOT/FRA/ORD-13/47.

[22]Joint Committee on Standards for Educational Evaluation (2011). The Program Evaluation Standards. Los Angeles, CA: Sage.

From 2007 to 2017, the Evaluation Team received an internal review from Dr. Michael Zuschlag, a research methodologist at Volpe who was not a member of either the C^3RS Implementation Team or the Evaluation Team. Starting with the midterm analyses in 2012, Dr. Wayne Nelson served as an external reviewer in addition to Dr. Zuschlag. Dr. Nelson is a nationally recognized expert in the quantitative analysis of occurrence data. Specifically, Drs. Nelson and Zuschlag focused on:

- Suitable preparation of data for analysis.
- Selection of the specific analyses.
- Assumption testing to determine appropriate statistical tests.
- Interpretation of the results of the analyses.
- Consideration of alternative approaches.

The experts' reviews included face-to-face and conference call discussions with the Evaluation Team as well as inspection of the data, analyses, statistical testing, and written interpretation of the results. In written and verbal feedback, the experts confirmed the analyses and recommended additional and more suitable alternative analyses. This rigorous review process was performed with the quantitative analyses at each demonstration pilot site during midterm and final evaluations to ensure the most scientifically credible methods and results were presented in briefings, reports, and conference presentations.

3.7 Summary of Evaluation

This final report documents a summative and formative evaluation of the four C^3RS demonstration pilots. The evaluation questions are related to the implementation, impact, and sustainability. The purpose of this final report is to answer the evaluation questions by assessing multiple qualitative and quantitative measures. The detailed logic model for the evaluation is illustrated in Figure 8.

4. Evaluation Methods

To conduct an independent evaluation of the four C^{3}RS demonstration pilots, FRA RD&T assembled a team from the Volpe Center, who contracted with Jacobs and its subcontractor partner Syntek Technologies. The Evaluation Team completed its planning in 2006 and began collecting data as the first site started up in 2007. The Evaluation Team continued collecting data through 2016, after the fifth year of the fourth site's program; the team collected data for a total of 10 years.

4.1 Evaluation Phases

For each demonstration pilot the Evaluation Team broke the evaluation period into three phases: baseline, midterm, and final. Each phase consisted of data collection, analysis, preparation of a briefing presentation, feedback to the demonstration pilot site, and obtaining their interpretation of the findings. Table 4 shows the type of evaluation (implementation, impact, and sustainability) conducted during each phase. Implementation Evaluation was a focus at baseline and midterm; Impact and Sustainability Evaluation were the focus during the final phase. Explanations of impact and sustainability drew on data throughout the project because factors related to a program's ability to cause change and its staying power could express themselves at any time.

Table 4. C^{3}RS Phase vs. Evaluation Type

	C^{3}RS Evaluation Phase		
Evaluation	Baseline	Midterm	Final
Implementation	✓	✓	
Impact			✓
Sustainability			✓

4.2 Evaluation Methods

Figure 13 provides a schematic view of the data and possible comparisons that comprised the research design—a non-randomized quasi-experimental method depicted in three dimensions:

1. Demonstration sites are shown on the X axis (left to right).
2. Data sources are listed on the Y axis (top to bottom).
3. Evaluation phases are shown on the Z axis (front to back).

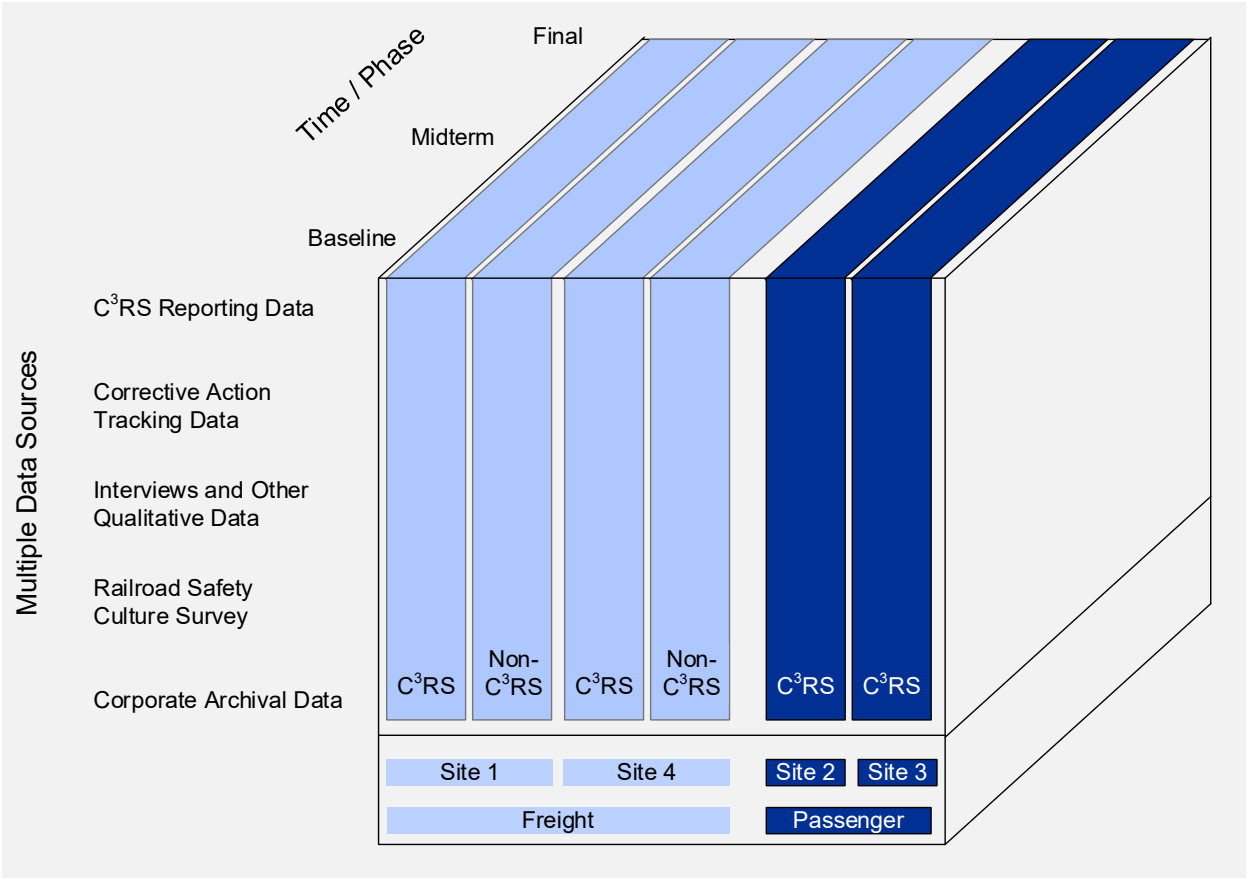


Figure 13. Schematic View of Evaluation Methodology

One way to look at Figure 13 is as a statement of the relationships that allowed the C^{3}RS program theory to be tested in the final evaluation. The methodology depicted in Figure 13 shows the multiple types of data assessed at different points in time at each of the four sites. The overall method represented in Figure 13 can also apply to more fine-grained questions about C^{3}RS's Theory of Action. For instance, did individual railroads improve safety and safety culture over time in the presence of C^{3}RS? What aspects of implementation were most important? Did certain impacts depend on certain implementation activities and implementation strength? Did sustainability depend on particular configurations of implementation and impact? The Evaluation Team could answer all these questions about C^{3}RS's Theory of Action because of the logic in Figure 13**,** which appeared in the evaluation plan.

Table 5 summarizes the methods of data collection, the types of data, and the role of each in answering the evaluation questions. The Evaluation Team collected data with each method repeatedly or continuously to assess changes associated with the operation of C^{3}RS. The data collection and analysis methods used in the final evaluation are described in the rest of this section.

Table 5. Methods of Evaluation and Their Relation to Logic Model Elements

Method of Data Collection	Type	Assessment
C^3RS Reporting Data	Quantitative	Implementation
Corrective Action Tracking Data	Qualitative	Implementation
Interviews and Other Qualitative Data (Implementation Interviews, Phased Interviews, Project Records, and Field Notes)	Qualitative	Implementation Impact Sustainability
Railroad Safety Culture Survey	Quantitative	Impact
Corporate Archival Data	Quantitative	Impact

Figure 14 illustrates the timeline for the evaluation methods superimposed on the schedule for each demonstration site. (In the figure, the timeline for each site is shown in parallel. The sites actually started in different years, but with exact dates disguised to protect site identities.) Note that the interview data extended past the 5-year demonstration period to collect information on sustainability.

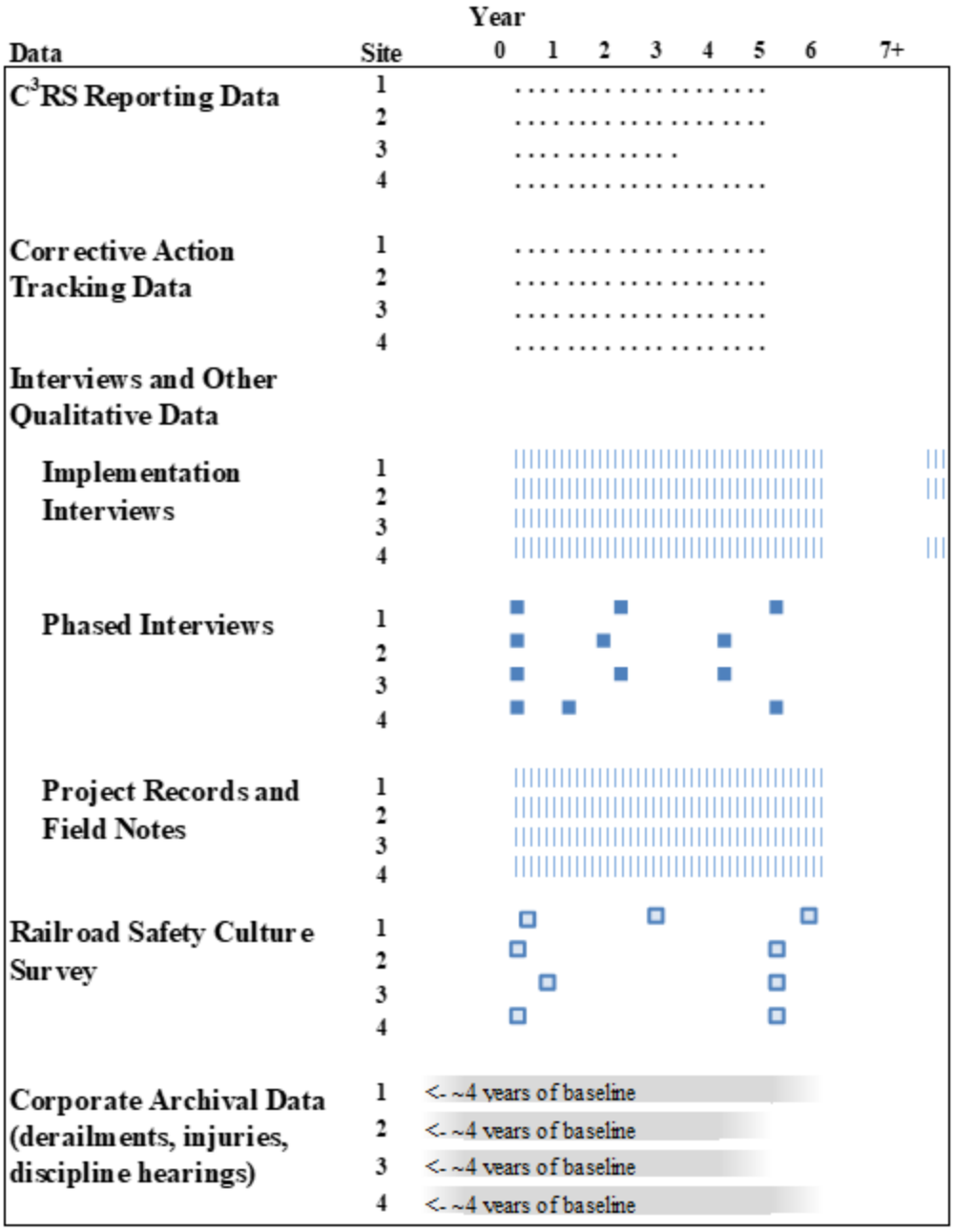


Figure 14. Data Collected for the C[3]RS Evaluation[23]

4.3 Confidentiality

The railroad carriers considered any data provided for evaluative purposes to the Evaluation Team as company confidential. Each carrier requested a non-disclosure agreement (NDA) with the Volpe Center that limited distribution of any data collected and findings summarized during

[23] Site 3 did not participate in the C[3]RS Survey. However, they did provide results from a similar safety culture survey to the Evaluation Team.

the evaluation. While the Volpe Center maintained editorial control over reports to be made public, the demonstration pilots reviewed drafts of reports and offered perspectives in advance of publication. Their ideas were welcomed and included in the final drafts of reports when appropriate.

In addition, the Evaluation Team contractor developed a data protection plan, which the Volpe Center carefully reviewed and approved. Highlights included:

- The Volpe Center operated under an NDA with each participating railroad.
- The Volpe Center included similar data protection clauses for all contractors working on the project.
- All evaluators received CIPSEA training by BTS.[24]
- Files were stored "off-network" in encrypted drives, encrypted external hard drives, and/or a GSA approved safe.
- Interview notes did not contain exact dates or personally identifying information.
- All project records used a notation marking them as company confidential, "For Official Use Only (FOUO). This document may be exempt from mandatory disclosure under FOIA. Company Confidential."

4.4 C^{3}RS Reporting Data

4.4.1 Data Collection

Monthly Reporting Frequency Data: An eligible reporter submitted each "C^{3}RS report" at a demonstration site after they observed a close call event. Recall that a close call event does not have to involve an actual unsafe event but can refer to an unsafe situation that a reporter observed.

The frequency of employees submitting reports per month was collected for each of the four sites. The Third-Party (BTS or NASA) provided this monthly reporting rate data to the Evaluation Team via the C^{3}RS Implementation Team. The Implementation Team also provided the eligible number of reporters to allow for normalization of the data to compare it across sites and over time.

C^{3}RS-Analyzed-Case Data: After the Third-Party received individual C^{3}RS reports from employees; its analysts created a report-record containing a redacted compilation of all reports on the same event. Before sending the report-record to the PRT, the Third-Party removed the following to ensure the reporter's confidentiality: details on the event, its location, and any personally identifying information such as the reporter's name. Then the PRT performed MCIA analysis on each report-record and recorded it in an MCIA tool as a C^{3}RS-analyzed-case. The PRT determined the contributing factors (CF), root causes, and corrective actions during their MCIA. The PRTs provided summary C^{3}RS-analyzed-case information from the MCIA tool's database to the Evaluation Team. As a matter of convenience, they provided the data to the

[24] Title V – Confidential Information Protection and Statistical Efficiency. 116 Stat. 2962. Public Law 107–347 – Dec. 17 2002. Available at https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/inforeg/proposed_cispea_guidance.pdf

Evaluation Team via the Third-Party for three of the sites. The fourth PRT provided a summary of their C^{3}RS-analyzed-case data directly to the Evaluation Team.

The C^{3}RS-analyzed-case data fields from three of the sites appear in Table 6.

Table 6. C^{3}RS-Analyzed-Case Data Provided to the Evaluation Team

Data Field	Description
Case ID	A numerical identifier of the C^{3}RS-analyzed-case.
6.4/CC	The type of close call • 6.4 = close call event where the reporter was given indemnity for providing information and was covered under the discipline protection (IMOU Section 6.4), for example, a minor derailment or run-through switch. In return the reporter also informed the railroad about the event. • CC = all other close calls.
Close call category	The category of the close call event, as identified by the Third-Party (e.g., Excess Speed, Yard – switch, Yard – derail, etc.).
Close call date	The date the close call event occurred (month and year only).
PRT date	The date the PRT created the C^{3}RS-analyzed-case and entered the results into their records (month and year only).
Contributing factors	The PRT assigned contributing factors from a pre-determined list that resided in the MCIA software. (The list of contributing factors was the same for all PRTs.)
Root cause	The main cause identified by the PRT (this was narrative text, completed by the PRT during analysis).
Corrective action	Brief notes about the corrective action the PRT considered.

4.4.2 Data Analysis

Monthly Reporting Frequency Data: The Evaluation Team calculated the number of C^{3}RS reports per month per 100 eligible reporters for each railroad throughout the evaluation period. Monthly, they graphed frequency rates to show trends (month to month). Then they statistically analyzed the reporting data using the same strategies used to analyze occurrence data from the corporate archival data (see Section 4.9).

For report purposes, the Evaluation Team calculated "trends" using the number of reports submitted to the Third-Party. It did not refer to the population of every close call that occurred at the railroad. For example, the Evaluation Team could only count the number of times a site's employees chose to report to C^{3}RS about speeding, not the total number of times speeding actually occurred.

C^{3}RS-Analyzed-Case Data: Using the C^{3}RS-analyzed-case information, the Evaluation Team first consolidated similar close call categories. For instance, they put multiple "excess speed" categories into a single overall "excess speed" category. They counted the numbers of C^{3}RS-analyzed-cases in each close call category and calculated percentages. Across sites, the Evaluation Team determined the most common close call categories.

4.5 Corrective Action Tracking Data

4.5.1 Data Collection

As mentioned earlier, an important aspect of the predicted success of C^{3}RS was to implement corrective actions to address issues raised by reporters. Implemented corrective actions signaled to reporters that they were being heard as well as the implementation of a mitigation to address the concerns raised. To assess corrective action implementation, the Evaluation Team requested and obtained corrective action tracking data from each PRT. A significant amount data was missing in what was provided that varied from PRT to PRT. Some PRT and Support Team corrective action records provided clarity about the underlying close call category of the C^{3}RS-analyzed-case and the corrective action to address it, others did not. Some railroads kept track of the corrective action implementation data and the department or person responsible for implementation while others did not. At times, it was not clear if the corrective action had only been recommended or was also implemented. The Evaluation Team addressed these missing data in the analysis by including questions about corrective action implementation in the interviews with PRT and Support Team members, and found additional data in newsletters and railroad presentations. To avoid double counting, they excluded interviews and field note information provided in the corrective action data set from the interview and field note data set.

The Evaluation Team predicted that effective corrective actions would have a positive impact on safety; therefore, they sought evidence of the carrier's assessment of effectiveness. However, little evidence existed as supplied by any of the four sites to suggest that they were assessing the effectiveness of any implemented corrective actions.

4.5.2 Data Analysis

The corrective action data were used for multiple purposes by the Evaluation Team.

- **Purpose 1: Determining which corrective actions were implemented versus not.** By examining the corrective action data, the Evaluation Team determined if a site used corrective action recommendations from the PRT to make actual changes at their site. This was important because implementation of actions could be difficult and was not guaranteed to occur (see Figure 9, Theory of Action Logic Model item 13).
- **Purpose 2: Determining which safety impacts might show observable changes.** The close call categories of C^{3}RS-analyzed-cases, e.g., excess speed or derailments, leading to specific corrective actions suggested the impact measure that could be impacted and built justification for further quantitative analysis (see Table 3, Mixed Methods: Building). For example, derailment corrective actions might reduce derailments. The number of corrective actions marked as implemented in the corrective actions tracking data suggested the strength of the implementation for each close call category. For missing data, for instance, if the close call category was not provided, the Evaluation Team inferred the appropriate category based on the corrective action description. For example, a corrective action to improve communication about the removal of derails was likely created to help reduce derailments. This analysis provided clues to the Evaluation Team about the corporate safety data most likely to be impacted (see Figure 9, Theory of Action Logic Model item 17).

- **Purpose 3: Determining challenges experienced in corrective action implementation and tracking.** The Evaluation Team analyzed missing data and the challenges the railroads had in providing consistent and complete corrective action tracking data. This analysis shed light on problems the demonstration pilot sites had in implementing and tracking corrective actions (Figure 9, Theory of Action Logic Model items 11, 12, and 13).

4.6 Interviews and Other Qualitative Data

The evaluation included four sources of qualitative data:

- **Phased Interviews:** Conducted face-to-face interviews with managers and workers at each site during each phase of the evaluation (e.g., baseline, midterm, and final) to provide qualitative data concerning perceptions of the implementation, sustainability, and impacts (Patton, 2002).
- **Implementation Interviews:** Conducted interviews periodically throughout the evaluation period to provide qualitative data concerning perceptions of the ongoing supports and barriers to implementation. The team conducted these interviews over the phone and occasionally face-to-face.
- **Project Records:** Reviewed documents created by C^{3}RS program participants to confirm other data collected, such as interviews, and to fill in missing data on corrective actions. Records included newsletters, brochures, email communication, and presentations.
- **Field Notes:** Recorded field notes by the Evaluation Team during conference calls and meetings. Meetings included monthly check-in conference calls conducted by the C^{3}RS Implementation Team with each demonstration pilot site, C^{3}RS User Group conference presentations, impressions recorded after each site visit, and informal discussions.

For all interviews, the Evaluation Team provided confidentiality by excluding personally identifying information from the notes, marking documents with a For Official Use Only (FOUO) notation, encrypting electronic files, and shredding paper notes. They conducted face-to-face and phone interviews typically in pairs, with one primary interviewer and one note taker. Both members of the interview team agreed on the final sets of notes used as data.

The Evaluation Team completed an Office of Management and Budget (OMB) application to collect interview data and received all the necessary approvals.[25]

4.6.1 Data Collection

4.6.1.1 Phased Interviews

To obtain an outsider's point of view, phased interviewees were primarily not members of the PRT or Support Teams. Workers eligible to report close calls but not on the PRT were interviewed. However, manager interviewees included individuals who were not affiliated with the program as well as some who were affiliated either by being on the PRT or Support Team because the population of potential interviewees was much smaller than for labor. The

[25] OMB CONTROL NUMBER: 2130-0574

Evaluation Team requested that interviewees include a wide range of tenures, as well as supporters of C^{3}RS and skeptics.

The face-to-face, semi-structured interviews occurred at the demonstration pilot site and lasted approximately 30 minutes. With significant employee schedule variability at a railroad, operating a 24/7 operation was typical. At each site, PRT members who were managers generally selected and coordinated employees for interviews who were available and met the Evaluation Team's criteria. Interviewees were eligible reporters who were at work on the days the Evaluation Team visited the site for interviews. Interviewees were sometimes selected ahead of time, but because of work schedule uncertainties it was not uncommon for managers to assign interviewees based on the flow of people coming in and out of the interview location. At each site, 10 to 15 interviews were requested per phase, typically spread over two days. The Evaluation Team requested one respondent per interview; on occasion, the carrier sent more than one person at a time. On these occasions, they conducted group interviews as it was the only practical way to obtain the data.

The phased interviews included these main topics of discussion:

- **C^{3}RS implementation strengths and weaknesses:** Questions about the acceptance of C^{3}RS, implementation supports and barriers, and how to improve implementation.
- **C^{3}RS impacts:** Questions about the categories of corrective actions that C^{3}RS was implementing and their potential impact on corporate measures and safety culture.
- **Possible confounds to any observed impacts:** Questions about other safety programs that might be alternative explanations for any observed impacts.
- **Labor-management relations:** Questions about labor and management cooperation.

For the complete list of questions refer to Appendix B. List of Phased Site Interview Questions.

4.6.1.2 Implementation Interviews

The Evaluation Team conducted implementation interviews on an ongoing basis to obtain the views of all the stakeholders involved in the C^{3}RS program over time (Figure 14). Interviewees included: PRT members (e.g., labor, management, and FRA); the C^{3}RS Implementation Team; the Support Team and other senior managers; the C^{3}RS Third-Party (BTS and NASA, as applicable); and FRA RD&T HF and RRS. They asked interviewees about key events related to the implementation, barriers and supports to implementation, demonstration site/FRA/Third-Party functioning, impacts, and sustainability.

The Evaluation Team used email and phone calls to request interviews. Most interviews were with individuals, but some were with small groups. Occasionally, they interviewed the PRT as a group.

Interviews were semi-structured, conducted by phone, and lasted approximately 30 minutes. Implementation interviews included these main topics of discussion:

- **C^{3}RS implementation:** Questions about key events, barriers, and supports.
- **Implementation functioning:** Questions about the functioning of the PRT, Support Team, FRA, and Third-Party.
- **Impact of C^{3}RS on safety and safety culture:** Questions about what corrective actions might impact corporate safety measures.

- **Sustainability of C^{3}RS:** Questions about C^{3}RS's ability to maintain itself in the long run.

In some cases, the Evaluation Team omitted certain topics if they were not appropriate for the interviewee. For the complete list of questions see Appendix C. List of Implementation Interview Questions.

In addition to the formal 5-year pilot period, the Evaluation Team conducted post-pilot implementation interviews in three of the four sites a few years after the demonstration period, thus providing useful information on sustainability.

4.6.1.3 Other Qualitative Data (Project Records and Field Notes)

In addition to the interviews, five other sources of qualitative data existed: 1) field notes as handwritten accounts of direct observations and interactions with stakeholders; 2) notes from the Evaluation Team's participation in the monthly 30-minute check-in calls that took place between the Implementation Team and each site's PRT; 3) notes from the Evaluation Team's participation in C^{3}RS's yearly User Group meetings; 4) newsletters and other C^{3}RS documents provided by PRTs; and 5) numerous ad hoc communications that frequently took place between the Implementation Team and the Evaluation team.

These notes and records provided information on the effectiveness of the implementation and its impacts on the stakeholder organizations.

4.6.2 Data Analysis

The Evaluation Team needed a systematic analysis approach to appropriately analyze the large volume of qualitative data. They used a multiple-step process for this analysis:

1. To observe pattern and meaning in the qualitative data, the data were content analyzed and coded for themes. See "Coding" below.
2. Using the coded qualitative data, the C^{3}RS reporting data, and the corrective action data, the Evaluation Team checked to see if the site implemented the activities on Figure 3 as planned.
3. Using the coded data, factors were identified that explained the extent to which implementation was achieved and the relationships among those factors. That knowledge was expressed in an Ishikawa "fishbone" or "cause-and-effect" diagram (Ishikawa, 1982). See "Cause-and-Effect Diagram Development for Implementation."
4. A 4-point rubric provided a standard "measuring stick" for the implementation factors, which the Evaluation Team applied to assign a rating to each implementation factor (Step 3) in the fishbone diagrams for each site. See "Rating Implementation Factors."
5. Using the same method as in Step 3, another Ishikawa fishbone/cause-and-effect diagram was developed to explain sustainability across the industry. This depicted the causes, or factors, influencing the sustainability of C^{3}RS across the railroad industry as a whole. See "Industry Sustainability Cause-and-Effect Diagram."
6. The Evaluation Team applied a similar rubric process of assigning ratings to the industry sustainability diagrams. See "Rating Industry Sustainability Main Factors Using a Rubric."

The Evaluation Team used the site implementation ratings to understand relationships between implementation, impact, and sustainability.

The Evaluation Team used the industry sustainability ratings to understand where C^{3}RS seemed to be working well in the industry and where it was not.

4.6.2.1 Coding

The Evaluation Team coded the qualitative data using standard qualitative data analysis methods (Patton, 1987 and 2002). Atlas.ti software was used to organize the data and codes.[26] The purpose of this analysis was to prepare each demonstration pilot's phased feedback session. The Evaluation Team employed a systematic process:

1. For each single interview or field note, they created a document containing notes and observations. Each document was added to the database and categorized in three ways: railroad (e.g., UP, CP, NJT, Amtrak); interviewee class (e.g., labor, management, FRA); and data type (e.g., baseline phased interview, implementation interview, field notes).
2. They read each interview or field note document and broke it into "comments." A comment consisted of a few sentences from a quote on a single topic. As a starting point for code development, comments were assigned codes that reflected elements in the Theory of Action Logic Model (Figure 9). Examples of codes are: PRT meetings, labor-management relationships, and impact on safety. (See Appendix D. List of Qualitative Data Codes.) Quality control was done by periodically showing a selection of coded comments to a second team member to confirm consistent coding.
3. They created additional codes as needed, based on additional comments from interviews, field notes, project records, trip impression notes, and their developing understanding of the program.
4. For each site's phased feedback briefing the Evaluation Team used Atlas.ti to identify the most frequent codes. They reread and aggregated interview comments under the most frequently used codes. To complete the analysis, they compared these results to the trip impression notes compiled by the interviewers. The team members then worked together to create and interpret summaries of the interview results and prepared a briefing for the site.
5. As a validity check, during each demonstration site's phased briefing (e.g., during the UP midterm briefing to the PRT), they discussed the interview results with the site to gain their perspective on the accuracy and clarity of the data. As appropriate, the site's feedback was incorporated into future briefings and reports.
6. They then used the validated data for subsequent evaluation activities described in the following sections.

4.6.2.2 Cause-and-Effect Diagram Development for Implementation

[26]Atlas.ti helps researchers to extract meaning from narrative data by allowing them to impose content categories, to rearrange and revise the categories over the course of an analysis, to nest and overlap groupings, and to sort out multiple categories. Available at http://www.atlasti.com/

Based on data analysis of each site over its 5-year pilot period, Ishikawa cause-and-effect, also known as fishbone, diagrams were developed for both implementation and industry sustainability (Ishikawa, 1982). In a cause-and-effect diagram, a single outcome or effect is represented at the right tip of the arrow, the "effect box." In Figure 15, the effect box at the right is favorable/strong/successful implementation. Going from left to right are the potential "causes," or factors that relate to the effect, which look like the bones of a fish hence the term fishbone. With that analogy, bones connected to the spine represent the main factors influencing the outcome or effect. Inputs to each main factor connect to the appropriate "bone." These inputs are known as detailed factors (Figure 15). The diagram organizes and visualizes the potential causes of a given effect. Main factors are considered important if sufficient evidence exists that they were necessary for the implementation to proceed as planned.

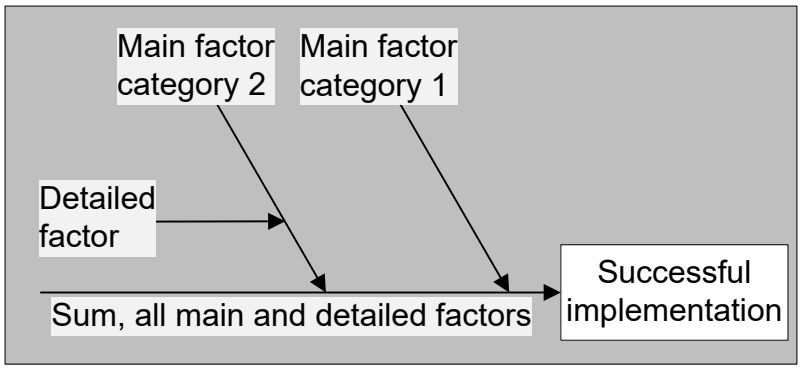


Figure 15. Structure of an Ishikawa Diagram

The Ishikawa diagram is flexible in that evaluators can use main factors and detailed factors that are meaningful to their setting. For instance, typical inputs in manufacturing include materials, manpower, machines, and technology while factors used in service marketing are surroundings, suppliers, systems, standard documentation, and skills.

To create the implementation fishbone/cause-and-effect diagram for C^3RS, the Evaluation Team applied following process:

1. The Evaluation Team began by reviewing the interview and other qualitative data and comparing the differences between sites with and without observed impact and sustained involvement in C^3RS after the demonstration. The sites were compared on different aspects of the C^3RS implementation. For example, when interviewees at a sustained site where impacts were observed talked about the helpfulness of strong senior management support, and interviewees at a site without sustainment talked about how management support was lacking, then "Systemwide Champion" became a relevant detailed factor. As a second example, when interviewees at a sustained site talked about how labor helped implement corrective actions and interviewees at a site without sustainment talked about labor not being involved in corrective actions, then "Labor Help Implement Actions" became a relevant detailed factor. In addition, the Evaluation Team reviewed C^3RS's Theory of Action models to augment the diagram (see Figure 8 and Figure 9).
2. The Evaluation Team created a "main factor" for each of the roles (e.g., FRA, Third-Party, Carrier Management, and Labor) to organize the "detailed factors" related to each role. To aid in understanding, they grouped these on the top of the diagram.

3. They created four more “main factors” to contain “detailed factors” that involved multiple stakeholders: Shared Responsibilities, Implementing Corrective Actions, Effective Dispute Resolution, and Perceived Value. They placed these on the bottom of the diagram. Shared responsibilities included internal and external stakeholders (carrier labor and management, FRA, national labor, and the Third-Party).
4. As a validity check of the usefulness of the approach, the diagrams for all the sites were not developed until feedback from the first site’s briefing on the implementation cause-and-effect diagram occurred. The feedback from the site during this briefing helped to further improve the validity of the method and also indicated that the approach was appropriate and useful.
5. Cause-and-effect fishbone diagrams were then improved incrementally based on feedback at subsequent site briefings.
6. After all four sites were done with their final evaluation, a single generic implementation fishbone diagram was created that could be applied to all sites during the final evaluation (see Figure 16). The definitions of each of the detailed factors follow Figure 16. (In the diagram, some factor names are shortened to enhance the visuals; full descriptive names are used in the definitions.) The effect box is also defined.

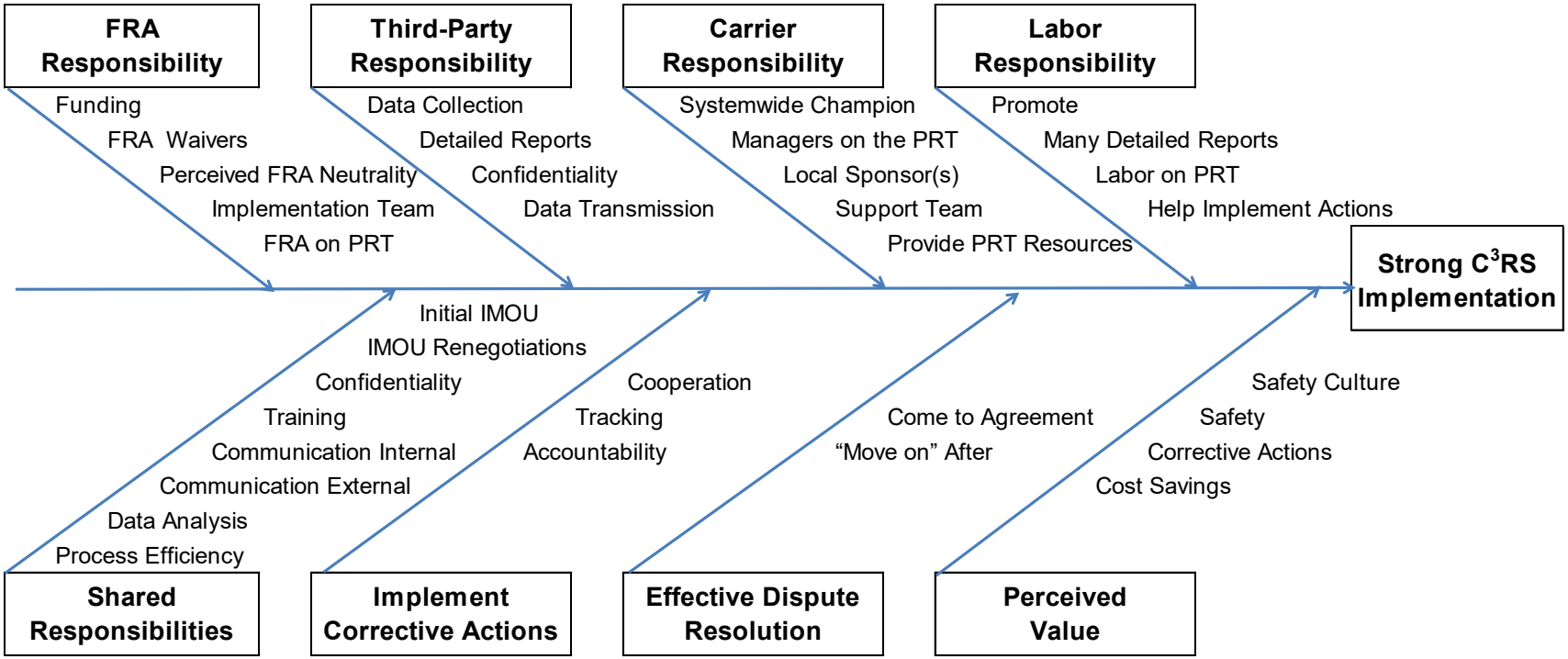


Figure 16. Implementation Ishikawa Cause-and-Effect Diagram for Demonstration Pilots

- FRA Responsibility
    - **Funding:** Extent to which FRA budget allocations supported C^3RS in recruiting new railroads, program management, implementation team training, and payment to the Third-Party.
    - **FRA-Granted Waivers:** Adequacy of waivers that legally permit participating railroads to refrain from taking disciplinary action in return for employees submitting C^3RS reports; and protection for railroads from fines imposed by FRA in the event that an employee was not disciplined.
    - **Perceived Neutrality of FRA:** Extent to which railroad labor and management perceived FRA to be unbiased with respect to disputes about C^3RS protections.
    - **Assistance from Implementation Team:** Extent to which the C^3RS Implementation Team's efforts were sufficient to roll out and operate the PRT, including obtaining reports, analyzing cases, and developing/implementing corrective actions.
    - **FRA on PRT:** Extent to which participation PRT meetings, by regional FRA staff, supported the analysis of cases and development of corrective actions.
- Third-Party Responsibility
    - **Data Collection:** Effectiveness of data collection by the Third-Party, including the amount and relevance of information collected and the usefulness of questions on the forms and during the phone callbacks with reporters.
    - **Detailed C^3RS Report Records to PRT:** Extent to which the PRT understood the C^3RS report-records sufficiently to develop an effective corrective action.
    - **Confidentiality:** Extent to which the Third-Party did not disclose information to the PRT or to the public that might identify the person who submitted a C^3RS report and protected the railroad from public knowledge of their close calls.
    - **Transmission of Report Records to the PRT:** Extent to which the Third-Party efficiently transmitted C^3RS report records to the PRT in a manner that supported effective action by the PRT.
- Carrier Responsibility
    - **Systemwide Champion:** Extent to which an influential senior manager was present and active in: 1) protecting and promoting C^3RS within the company and 2) exerting the authority necessary to implement corrective actions.
    - **Managers on the PRT:** Extent to which management participation on the PRT led to the development and implementation of local corrective actions.
    - **Local Sponsor(s):** Effectiveness of manager(s) whose responsibility was primarily within the boundaries of that part of the carrier implementing C^3RS. These managers might not be on the PRT, but were responsible for implementing local corrective actions.

- **Cross-Functional Senior Management (Support Team):** Effectiveness of the senior management team representing the functions that may need to be involved to implement corrective actions, including effectiveness of the team's communication with the PRT.
- **Provide Resources (for PRT):** Extent to which the carrier provided resources for PRT time and travel.

- Labor Responsibility
  - **Promote:** Extent to which PRT members promoted C^3RS, both formally in union meetings and informally with peers, and the extent to which PRT members were available and willing to answer questions from labor and management.
  - **Detailed and Numerous Reports:** Extent to which eligible reporters provided detailed and numerous C^3RS reports concerning a variety of close calls and did not just use the program to "get-out-of-jail."
  - **Labor on PRT:** Effectiveness of PRT labor members, e.g., effective participation in meetings, smooth rotation of members on the PRT, and "leaving hats off" (i.e., focused on the goals of C^3RS, not the goals of the group to which they belonged).
  - **Help with Corrective Actions:** Extent to which labor helped implement corrective actions, especially when local action by labor and management could lead to effective change.
- Shared Responsibilities – Includes FRA, carrier labor and management, national labor, and the Third-Party
  - **Initial IMOU:** Existence of an initial signed IMOU between labor, management, and FRA for that site.
  - **IMOU Renegotiations:** Effectiveness of the IMOU re-negotiation processes as needed.
  - **Confidentiality:** Extent to which all stakeholders had a common understanding of C^3RS confidentiality terms and the degree to which they protected that confidentiality.
  - **Training:** Adequacy of training for PRT concerning the provisions of C^3RS, MCIA process, usage of tools, and roles and responsibilities.
  - **Communication between C^3RS Internal Stakeholders:** Effectiveness of communication between the PRT and the Support Team.
  - **Communication to Workforce and Management External to C^3RS:** Effectiveness of outreach to workforce and management concerning the intent, usage, and achievements of C^3RS.
  - **Data Analysis:** Extent to which the PRT's data analysis revealed root and contributing causes, trends, and cases which should be dealt with as a group of related issues and safety priorities for action.
  - **Process Efficiency:** Extent to which stakeholders worked together to make C^3RS processes more efficient.

- Ability to Implement Corrective Actions
    - **Cooperation:** Extent to which relevant parties were able to cooperatively develop and implement corrective actions. "Relevant parties" included PRT labor and management, local management, the Support Team, and senior management, as needed.
    - **Tracking:** Extent to which the participants tracked and updated the status of corrective actions.
    - **Accountability:** Extent to which specific people were clearly assigned responsibility for implementing each corrective action and held accountable.
- Effective Dispute Resolution
    - **Come to Agreement:** Extent to which participants in a dispute were able to either resolve the dispute or "agree to disagree" in a timely manner.
    - **"Move on" After:** Extent to which the PRT continued to function after negotiating a dispute.
- Perceived Value
    - **Safety Culture:** Extent to which C^{3}RS was perceived as contributing to safety culture improvements, whether or not there was empirical evidence to justify such a claim.
    - **Safety:** Extent to which C^{3}RS was perceived as contributing to safety improvements, whether or not there was empirical evidence to justify such a claim.
    - **Corrective Actions:** Extent to which corrective actions that were implemented were perceived to be effective, regardless of whether empirical data existed for their effectiveness.
    - **Cost Savings:** Extent to which C^{3}RS and its corrective actions resulted in perceived cost savings (or its time equivalent), whether or not there was empirical evidence to justify such a claim.
- **Strong Implementation** (the effect box): Extent to which C^{3}RS implementation was completed as planned and was favorable/strong across a majority of factors.

4.6.2.3 Rating Implementation Factors for Demonstration Sites Using a Rubric

The Implementation Cause-and-Effect Diagram for Demonstration Pilots, as shown in Figure 16, depicts important factors for C^{3}RS implementation and how they relate to each other. The next step was to estimate for each demonstration the extent to which each main factor and detailed factor were implemented as planned and whether evidence existed to suggest it was or was not sufficient. This allowed implementation similarities and differences across sites to be explained. After impact analysis was completed, the implementation factor ratings also enabled an assessment of which implementation factors might relate to impacts. The purpose was to help railroads learn how to keep their C^{3}RS programs strong and useful.

To assign ratings to the implementation factors, the Evaluation Team applied the following process:

1. Develop a rubric – a table containing values and definitions for a set of ratings that could be consistently assigned to implementation factors.
2. Assign ratings from the rubric to the implementation factors at the four sites.
3. Compare results across sites.

The Evaluation Team developed a rubric that ensured a standard set of ratings would be applied to each implementation factor at each site, based on Davidson's "Evaluation Methodology Basics" (Davidson, 2005). The rubric comprised a table containing four evaluative ratings (poor, fair, good, very good) and objective, observable operational definitions for those ratings (Table 7). The number of categories and their definitions were chosen to reflect the Evaluation Team's judgment of the widest range of meaningful assessments while avoiding the threat of false precision. The ratings were applied in an "absolute" sense, reflecting "ideal" implementation, as opposed to a relative sense of how the sites performed relative to each other. The absolute standard allowed the Evaluation Team to identify factors that were challenging to all sites, even those with better implementation.

Table 7. Rubric for Rating C^{3}RS Implementation Factors

Qualitative Rating	Quantitative Rating	Definition
Very Good	4	Execution of implementation factor is clearly exemplary, could not have been substantially better. Any gaps or weaknesses are not significant with respect to C^{3}RS operations and are managed effectively.
Good	3	Execution of implementation factor is functional and adequate. A few weaknesses may exist, but none are considered overly problematic.
Fair	2	Execution of implementation factor is inconsistent and/or has multiple weaknesses. It does not adequately support C^{3}RS operations, but did not pose a major threat on its own.
Poor	1	Execution of implementation factor has numerous weaknesses, some of which pose a serious threat to C^{3}RS operations.

Given that the ratings were qualitative, assigning a number to each rating is worth explaining. The rubric made up an ordinal scale (i.e., it contained rank order information), but provided no information about the degree of difference between the ratings. For instance, the Evaluation Team knows that "very good" is better than "good," but they may have no information about how much better the one rating is from another. Interval scales do contain information on the degree of difference among items. Dates are a good example: March 4th is the fourth day of the

month, and that two days intervene between the 1st and the 4th.[27] Despite this technical difference between ordinal and interval scales, there is precedent in the scientific literature for treating ordinal scales as having "interval-like" properties, thus, with appropriate cautions of interpretation, allowing arithmetic operations to be performed (Petty, 2016).

After creating the rubric, the Evaluation Team assigned ratings to the detailed implementation factors for each site and applied a multiple-step approach for this:

1. A single member of the team made a first-pass of assigning a rating to each detailed factor for each site using the absolute standard in terms of whether the demonstration was performing the activity (1) poorly, (2) fairly, (3) good, or (4) very good. Each factor had a definition, describing what was supposed to be done, which was considered when assigning the rating. Information presented in the final briefing for each demonstration was the basis of the rating. If the final briefing did not contain sufficient information to assign a rating, then original qualitative data (e.g., interviews, field notes) were consulted.
2. A second member of the team independently reviewed the ratings, repeated Step 1, and identified disagreements. The two members discussed every rating, reviewed the qualitative data as needed, and reached consensus on each one.
3. A third member of the team then reviewed the ratings from Steps 1 and 2. They discussed each rating and reached a consensus, drawing on the same information outlined in Step 1 if questions arose.
4. While doing Step 3, the team compared the ratings among the sites. To do this, the team examined the relative activity of the sites to ensure that the sites with more activity on a detailed factor had higher ratings than the sites with less activity, and made adjustments as needed.
5. After assigning each demonstration pilot's detailed factor a rating, the Evaluation Team calculated an average for each main factor using the ratings of the detailed factors.
6. The team created fishbone/cause-and-effect diagrams containing detailed factors ratings for each site.
7. For each site, the team also created a single table with the average main factors ratings.

The result of following the steps outlined above was, for each site, 1) creating a fishbone diagram with ratings for all detailed factors and 2) creating a table showing the average ratings for the main factors. The team used these figures to compare similarities and differences across the sites using an absolute standard.

For example, the detailed factor "Carrier Responsibility: Local Sponsor" was defined as: "Effectiveness of manager(s) whose responsibility is primarily within the boundaries of that part of the carrier that is implementing C^3RS. These managers might not be on the PRT, but are responsible for implementing local corrective actions." For Site 1, the Evaluation Team assigned a rating of "good." The initial local sponsor at Site 1 was very enthusiastic at the beginning of the demonstration. He attended some PRT meetings to discuss corrective actions with the PRT

[27] Wikipedia. Ranking.

and ensured that resources were devoted to implementing corrective actions. Later, as turnover occurred, local sponsors at Site 1 had less enthusiasm but still provided adequate support to keep the program implemented as planned during the demonstration. The Evaluation Team determined that this detailed implementation factor should be rated as "good" because it fit the definition, "Execution of implementation factor is functional and adequate. A few weaknesses may exist, but none are considered overly problematic."

4.6.2.4 Summary Demonstration Pilot Cause-and-Effect Diagram

The last step of the analysis was to assign summary ratings to the "effect boxes" at the right side of the implementation cause-and-effect fishbone for each site (Figure 16). The Evaluation Team decided to have a two-level "head" (effect box): one for impact and one for sustainability. For this summary, in addition to the qualitative results for each demonstration, the quantitative impact results for each demonstration are included. The quantitative impact results were added to portray a summary fishbone diagram for each site that included implementation, impact, and sustainability. This allowed more useful comparisons across the four sites.

The evaluation team used only two ratings for impact and sustainment. Concerning impact, either the Evaluation Team observed operationally substantial results or they did not. See Sections 5.4 and 5.5 for a description of the methods for obtaining the impact results. The Evaluation Team made a summary judgement based on all the available quantitative impact data. Concerning sustainment, either the demonstration continued with C^3RS after the demonstration period or it did not.

The definitions of Impact ratings were as follows:

- Substantial Impact (4) – operationally substantial impact on safety.
- Little Impact (1) – little to no impact on safety and safety culture.

The definitions of Sustainability ratings were as follows:

- Continued (4) – the site decided to continue C^3RS and was still involved at the time of the writing of this report.
- Withdrew (1) – the site had left the C^3RS program as of the time of the writing of this report.

The Evaluation Team also used the qualitative data, especially the post-pilot interviews and field notes, to determine the reasons why each site continued or withdrew. The team wrote a summary of those reasons and compared them across sites. The purpose of this was to explore what issues impact sustainability long-term after the demonstration.

4.6.2.5 Industry Sustainability Cause-and-Effect Diagram

In addition to the implementation fishbone/cause-and-effect diagram in Figure 16, the Evaluation Team developed a second fishbone diagram concerning sustainability at the industry level. The Evaluation Team collected much less data on industry sustainability than implementation at the demonstration sites. However, the purpose of the simpler industry sustainability diagram was to explore factors, outside the demonstration pilot sites, that influenced the ability of C^3RS to continue after the demonstration period and expand across the industry.

To create an industry sustainability cause-and-effect diagram, the Evaluation Team applied the following process:

1. The Evaluation Team began assessing the interview and other qualitative data about the sustainability of C^{3}RS in the industry. This included some interviews with FRA HPD about the growth of C^{3}RS in the industry, beyond the demonstration pilots.
2. They added a "main factor" for each of the stakeholders external to carriers to the top of the sustainability cause-and effect diagram (e.g., FRA, NASA, national labor).
3. They created three more "main factors" along the bottom of the model: C^{3}RS visibility and industry environment.
4. This cause-and-effect diagram was not created until the end of the evaluation. Therefore, the Evaluation Team did not always have the opportunity to discuss it with all represented stakeholders.

The Industry Sustainability Ishikawa Cause-and-Effect Fishbone Diagram is shown in Figure 17. The definitions of each of the factors follow the figure.

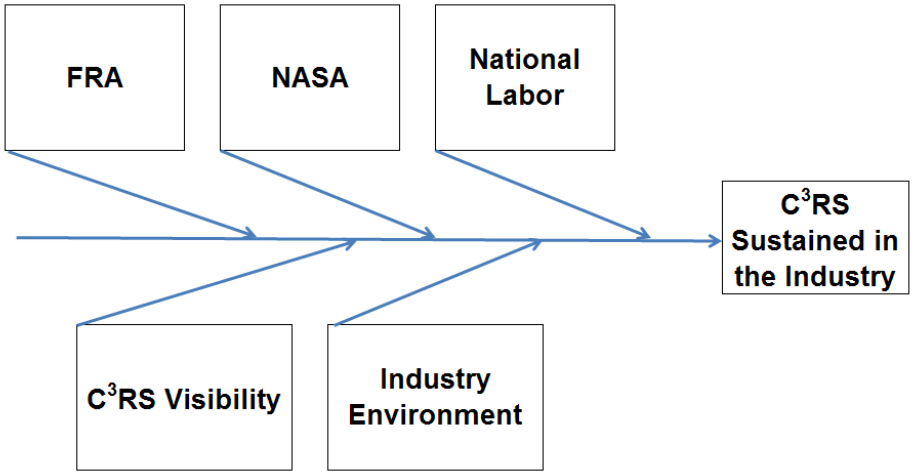


Figure 17. Industry Sustainability Ishikawa Cause-and-Effect Diagram

- **FRA:** Extent to which FRA supported the C^{3}RS program for the industry.
- **NASA:** Extent to which NASA supported C^{3}RS.
- **National Labor:** Extent to which national labor supported C^{3}RS.
- **C^{3}RS Visibility:** Extent to which the visibility of C^{3}RS sufficiently affected the combined motivations of the industry, the FRA, and Congress to a level that would support the continuation of C3RS in the railroad industry. In turn, the extent to which increased participation by carriers motivated more railroads to join and FRA to continue to support and fund C^{3}RS.
- **Industry Environment:** Extent to which events and circumstances in the railroad industry motivated participation in C^{3}RS.

4.6.2.6 Rating Industry Sustainability Main Factors Using a Rubric

The sustainability fishbone diagram depicts the factors important for C^{3}RS sustainability in the railroad industry. The next step was to estimate the extent to which each factor supported sustainability and whether there was evidence to suggest if it was or was not sufficient. This analysis concerned post-demonstration sustainability interviews, field notes, and project records. There was much less data on industry sustainability than for site implementation. The purpose of this analysis was to help FRA and railroad industry determine how to keep the C^{3}RS program strong and useful. The Evaluation Team used a similar rubric as was used for implementation with the evaluative ratings (e.g., poor, fair, good, very good) and objective, observable operational definitions for those criteria (Table 8).

Table 8. Rubric for Rating C^{3}RS Industry Sustainability Factors

Qualitative Rating	Quantitative Rating	Definition
Very Good	4	Execution of sustainability factor is clearly exemplary, could not have been substantially better. Any gaps or weaknesses are not significant with respect to C^{3}RS sustainability in the railroad industry and are managed effectively.
Good	3	Execution of sustainability factor is functional and adequate. A few weaknesses may exist, but none are considered overly problematic.
Fair	2	Execution of sustainability factor is inconsistent and/or has multiple weaknesses. It does not adequately support C^{3}RS sustainability in the railroad industry, but did not pose a major threat on its own.
Poor	1	Execution of sustainability factor has numerous weaknesses, some of which pose a serious threat to C^{3}RS sustainability in the industry.

After creating the rubric, the next step was to assign ratings to the main sustainability factors for each site using a multiple-step approach:

1. A single member of the team made a first-pass of assigning a rating of each main factor for each site using the absolute standard in terms of whether the demonstration was performing the activity (1) poorly, (2) fairly, (3) good, or (4) very good. They used information presented in post-pilot interviews and field notes as the basis of the rating.
2. A second member of the team independently reviewed the ratings, repeated Step 1, and identified disagreements. The two members discussed every rating, reviewed the qualitative data as needed, and reached consensus on each.
3. A third member of the team then reviewed the ratings from Steps 1 and 2. They all discussed each rating and reached a consensus, drawing upon the same information outlined in Step 1 if questions arose.

4.7 Railroad Safety Culture Survey

Forced-choice multi-scale perception surveys from multiple administrations provided data for evaluating the impact on safety culture as well as other data about C^3RS. There were two different survey forms, one for management, and one for labor respondents.

They administered the C^3RS survey at only three of the four sites. Site 3 had a parallel safety culture survey effort underway. The C^3RS survey was not administered at Site 3 due to concerns about low response rates and survey fatigue. Adding an additional survey would constitute a high response burden that could lower response rates. The Evaluation Team received summary reports of Site 3's safety culture survey effort.

4.7.1 Survey Administration

BTS prepared the survey forms and collected the survey data at the three sites. BTS also completed the Office of Management and Budget (OMB) approval process for the survey.[28]

4.7.1.1 Administration Times

The original evaluation plan called for three administrations of the survey at the three sites: baseline, midterm, and final. At midterm, only one site actually administered the survey as planned. After discussions with the PRTs, the Evaluation Team found that Sites 2 and 4 were having internal labor-management issues unrelated to C^3RS, which might influence the survey scores threatening their validity. Due to these concerns they decided to wait until the final evaluation phase to administer the survey at those two sites for the second time. Because only Site 1 completed a midterm survey, the Evaluation Team decided to only compare baseline and final survey responses for this report.

4.7.1.2 Survey Preparation

Each time they administered the survey, the Evaluation Team provided BTS with a codebook containing all questions and the forced-choices. BTS used the codebook to prepare the fixed-choice survey forms using eListen software. BTS also contacted the PRT to plan the times and locations for each administration.

4.7.1.3 On-Site Procedure

Each PRT was primarily responsible for on-site proctoring for the three sites that used the safety culture survey. This involvement allowed the PRT to pass out program literature and answer questions at the time of survey administration. To ensure respondent confidence in the anonymity of responding to the survey: 1) PRT members stayed away from respondents while they filled out the survey; 2) respondents were not asked for any identification; and 3) respondents were asked to put their surveys in a sealed envelope before returning them. This procedure was followed at all three sites. The PRT involvement provided workforce credibility to the survey and encouraged completion. BTS also provided proctors.

[28] The Paperwork Reduction Act of 1980, and as amended in 1995, requires Federal agencies to obtain approval from OMB each time they propose to collect or sponsor, even under a contract or other agreement, the collection of identical information from more than nine respondents.

BTS coordinated with the PRT and the manager of a given location to obtain PRT members to proctor the survey. BTS personnel traveled to each site to distribute and collect the surveys. BTS worked closely with each PRT to administer the survey and maximize the response rate at each demonstration pilot site. BTS sent two team members to each survey administration to accommodate multiple shifts at multiple locations over several days. (The exception to this process was the baseline data collection at one site. There, BTS worked with the site to prepare for data collection but did not have people present during data collection.)

On the day of administration for each site, PRT and BTS survey proctors were stationed in proximity to where employees initiated their workday. For example, at one site the survey was given in a conference room that the employees would walk by on their way to collect the paperwork for their shift. At another site, they stationed proctors near the locker room. Employees in some locations were called in early, while in other locations they completed the survey after their shift. At some locations, the railroad offered employees a snack to show their appreciation for participation. Some proctors traveled to collect surveys from employees at more remote locations. In cases where these tactics did not yield a large enough sample, a second round of survey data collection took place close in time to the first administration.

The survey took approximately 30 minutes for employees and managers to complete. Respondents were paid for their time by the railroad, and in one case FRA provided funding to the railroad for this purpose. Each respondent sealed their completed survey in an envelope and provided it to the PRT who mailed the completed surveys to BTS via overnight delivery.

4.7.1.4 Protection of Human Subjects

Survey respondents were informed of the procedures being used to protect them as human subjects via a cover letter that accompanied each survey (see Appendix E. Example Survey Cover Letter). The letter began by explaining the purpose of the C^{3}RS evaluation and how BTS would protect the identity of survey respondents. For example, survey respondents' names were not requested. BTS also assured respondents that the railroad would not receive the survey data, and that the Evaluation Team would only use the data for statistical purposes and to publish summary results. The letter was signed by labor and management representatives from the site.

4.7.1.5 Data Preparation

After administration, BTS processed the surveys at their office. BTS personnel used a Scantron® machine to scan each survey and in some cases manually entered the data. BTS personnel then checked for a match between the values on the paper survey and the scanned surveys. Handwritten comments to the open-ended question at the end were manually typed into the data set. The BTS confidentiality officer reviewed handwritten survey comments to remove any names or personally identifying information before it was entered. Once all the data were entered, BTS sent it to the Evaluation Team in an Excel file.

4.7.2 Respondents

The C^{3}RS Railroad Safety Culture Survey was completed by Sites 1, 2, and 4. Site 3 completed their own safety culture survey, using six scales with similar definitions that could be compared to the C^{3}RS survey. There were 3,521 participants in the C^{3}RS survey. The Evaluation Team chose to concentrate on the analysis of labor's changes over time, because: 1) management

respondents had higher baseline views, allowing for little change over time, and 2) there were limited management respondents at final.

The Evaluation Team did not receive information on the precise headcounts during the administration, so response rates could not be computed. However, response rates could be estimated by dividing the number of labor respondents by the approximate number of eligible reporters provided by the Implementation Team (Table 9).

Table 9. Number of Survey Respondents

		Total Number of Respondents		Approximate Response Rate		
		Labor	**Management**	**Site 1**	**Site 2**	**Site 4**
Phase	Baseline	1458	121	17%	47%	69%
	Final	1878	64	36%	51%	34%

The number of labor respondents included in each analysis was lower because not everyone completed all of the questions. An example from one analysis showed that at baseline 99% of respondents completed the analyzed questions across the sites. However, this varied among sites, with 91% of Site 1, 99% of Site 2, and only 54% of Site 4's respondents completing the applicable questions.

Looking at the demographics, some differences existed over time in the proportions of respondents for both tenure and location (yard vs. road). As a result, these were candidates for covariates in the analysis.

4.7.3 Survey Instrument

The C^{3}RS Railroad Safety Culture Survey included several sections: respondent demographics; railroad safety culture scales shown to relate to safety outcomes; impact and sustainability questions; C^{3}RS reporting questions; and an open-ended comment field. These survey sections are described below. All questions were fixed-choice, except for an open-ended item at the end of the instrument where respondents were provided with space to write any comment that they wanted to share.

4.7.3.1 Demographics

The first part of the C^{3}RS Railroad Safety Culture Survey asked each respondent to provide demographic information. The purpose of the demographic information was to allow examination of differences among types of respondents during data analysis to account for any observed biases in respondents that might influence survey scores. The demographic questions focused on:

- **Job Type:** labor or management.
- **Job Category:** trainman/conductor, switchman, locomotive engineer, hostler, yardmaster, dispatcher.
- **Location:** road, yard, both.
- **Shift:** mostly days, mostly afternoons, mostly nights, mostly variable shifts.

- **Gender:** male or female.
- **Age and Tenure:** grouped into time-span categories to keep responses anonymous.

4.7.3.2 Railroad Safety Culture Scales

To measure culture change, the Evaluation Team used scales from past Volpe studies to allow possible comparisons with other railroad studies occurring earlier or later. Only scales that had been used previously and had demonstrated validity and reliability in the industrial safety literature were used. The safety culture scales used in this evaluation are summarized in Table 10. These scales were fixed-choice on a range of 1 to 5; from strongly disagree to strongly agree. The Evaluation Team organized the scales into four categories: Organizational/ Managerial, Supervisor, Coworker, and Safe Behaviors. The site column in the table indicates which sites received each scale. Some sites did not receive certain scales due to the timing and need to shorten the length of the survey at certain sites.

Table 10. Railroad Safety Culture Survey Scales and Definitions

Scale Name in Survey	Description	No. of Items	Item Source	Scale Name in Literature	Site
Organizational/Managerial Scales					
Organizational Concern for Employees	The extent to which employees believe that the organization is concerned about their needs	3	Eisenberger, R., Huntington, R., Hutchison, S., et al. (1986)	Perceived Organizational Support	1, 2, 4
Labor-Management Relations	The extent to which employees believe that there is cooperation between labor and management in the organization	6	Dastmalchian, A., Blyton, P., & Adamson, R. (1989)	Industrial Relations Climate	1, 2, 4
Management Safety	Employees' perceptions of management's attitude toward safety	11	Mueller, L., Da Silva, N. Townsend, J.C., et al. (1999)	Organizational Value for Safety	1, 2, 4
Supervisor Scales					
Supervisor Fairness	The extent to which employees perceive that their direct supervisors treat them fairly	7	Niehoff, B.P., & Moorman, R.H. (1993)	Procedural Justice	1, 2, 4
Supervisor-Employee Relationships	The extent to which employees perceive the working relationship between themselves and supervisors to be strong	7	Wayne, S.J., Shore, L.M., & Liden, R.C. (1997)	Leader-Member Exchange	1, 2, 4

Scale Name in Survey	Description	No. of Items	Item Source	Scale Name in Literature	Site
Raising Concerns with Supervisors	The extent to which employees perceive that their supervisor is open to their raising safety concerns	7	Hofmann, D.A., & Morgeson, F.P. (1999)	Upward Communication	1, 2, 4
Coworker Scales					
Work-Safety Priorities	Employees' perceptions of how committed the organization is to safety over productivity	5	Mueller, L., et al. (1999)	Worker Involvement in Safety	1, 2, 4
Helping Behavior	The extent to which employees perceive that coworkers perform extra activities to help each other	4	Naumann, S.E., & Bennett, N. (2000) Organ, D., & Konovsky, M. (1989)	Procedural Justice Climate Altruism	1, 2, 4
Coworker Safety	The extent to which employees perceive that coworkers encourage each other to work safely	5	Mueller, L., Da Silva, N. Townsend, J.C., et al. (1999)	Approaching Others	1, 2, 4
Safe Behaviors Scales					
Safe Behaviors – Road Crews – Cab Red Zone	Self-reported extent to which road crews practice safe behaviors in the cab red zone	6	UP Code of Operating Rules, checked by UP subject-matter experts Hofmann, D.A, & Stetzer, A. (1996)	Safe Behaviors	1 (final only), 2,4
Safe Behaviors – Yard Crews – Switching	Self-reported extent to which yard crews practice safe behaviors during switching	6	UP Code of Operating Rules, checked by UP subject-matter experts Hofmann, D.A., & Stetzer, A. (1996)	Safe Behaviors	1 (final only), 2,4
Safe Behaviors – Dispatchers	Self-reported extent to which dispatchers practice safe behaviors during dispatching	6	Hofmann, D.A., & Stetzer, A. (1996)	Safe Behaviors	2
Safety Briefings	The extent to which safety briefings are part of doing business	1	Dedobbeleer, N., & Beland, F. (1991)	Worker Involvement in Safety	1 (final only), 2,4

4.7.3.3 Impact and Sustainability Questions

The survey also included questions written by the Evaluation Team to determine views of C^{3}RS's impact on safety and sustainability. The purpose of these questions was to determine these views from a wider audience than was reached by the interviews and to help answer the implementation and sustainability evaluation questions. These scales were fixed-choice on a range of 1 to 5; from strongly disagree to strongly agree.

Table 11. C^{3}RS Impact and Sustainability Survey Questions and Definitions

Question Name	Description	No. of Items	Site
Improved Relationship between My Railroad and the FRA	The extent to which employees perceive that C^{3}RS has improved the relationship between their railroad and the FRA	1	1, 2 (final only), 4
Increased Awareness of Safety Issues within Labor	The extent to which employees perceive that C^{3}RS has increased labor's awareness of safety issues	1	1, 2 (final only), 4
Increased Awareness of Safety Issues within Labor	The extent to which employees perceive that C^{3}RS has increased labor's awareness of safety issues	1	1, 2 (final only), 4
Resulted in Improved Safety	The extent to which employees perceive that C^{3}RS has improved safety	1	1, 2 (final only), 4
Sustainability	The extent to which employees perceive that C^{3}RS is likely to remain in operation at their railroad for the foreseeable future	1	1, 2

4.7.3.4 C^{3}RS Reporting Questions

The Evaluation Team wrote survey questions about C^{3}RS reporting. These questions were fixed-choice, but not on a five-point scale.

Table 12. C^{3}RS Reporting Survey Questions

Question	Choices	Site
Have you submitted a report to C^{3}RS?	Yes No I don't know	1, 2, 4

Question	Choices	Site
Do you personally know people who have submitted a report to C^{3}RS?	Yes No I don't know	1, 2 (final only), 4
Do you think you would recognize an incident that could be reported to C^{3}RS, if you saw one?	Yes No I don't know	1, 2 (final only), 4
Do you think that you would report an incident to C^{3}RS?	Yes No I don't know	1, 2 (final only), 4
What might keep you from reporting an incident to C^{3}RS? (select all that apply)	A. I am not familiar enough with the reporting procedure B. The reporting process is too much of a bother C. I don't trust the Bureau of Transportation Statistics to maintain confidentiality D. I don't think it would result in any changes E. I worry about being punished by management (Labor) E. I worry about being punished by my supervisors (Management) F. I do not know enough about C^{3}RS	1, 2, 4

4.7.3.5 Open Comments Question

Respondents were asked: "Please provide comments about safety at your site and the C^{3}RS program." Responses to this item were used in briefings to stakeholders at individual demonstration sites, but are not included in this report.

4.7.3.6 External Survey Data at One Site

Site 3 did not participate in the C^{3}RS Railroad Safety Culture Survey. However, they did complete their own culture survey and provided the Evaluation Team with summary reports. Some members of the Evaluation Team had worked with Site 3's survey before and identified six scales with similar definitions that could be compared to these C^{3}RS scales: Organizational Concern for Employees, Management Safety, Supervisor Fairness, Supervisor-Employee Relationships, Raising Concerns with Supervisors, and Coworker Safety.

4.7.4 Survey Analysis

The focus of the Evaluation Team's final analysis was both checking reliability of the scales and checking for changes in the scales over time to see if any sites improved safety culture.

4.7.4.1 Psychometric Performance

Even though the safety culture scales demonstrated reliability and validity from the literature, they were tested again by the Evaluation Team to ensure that they were reliable in the current setting. To aid in the railroads' understanding of the scale scores in feedback briefings, all the scales were transformed so that high scores consistently represented a "positive safety culture." Cronbach's alpha was conducted to test the reliability of the scales. In general, alpha scores of over 0.7 were considered adequate (Rosenthal and Rosnow, 1991).

The Evaluation Team also performed a factor analysis to determine if questions were loaded on the intended scales. As a result of the factor analysis, they made a couple of changes to the scales. They eliminated four questions from Raising Concerns with Supervisors due to poor factor loading. They split the Labor Management Relationships scale into two, keeping four questions, and moved the other two questions to the Management Safety scale. As Site 2 included shorter versions of Supervisor Fairness and Supervisor-Employee Relationships due to time constraints during their survey administration, they shortened those scales for all sites in the final analysis to be consistent.

4.7.4.2 Analysis for Changes Over Time

The final analysis focused on labor respondents. Management was excluded because 1) baseline values were high, leaving little room for improvement, and 2) management response rates at final were low allowing for few possible comparisons.

To look for changes over time in labor's perceptions in the Railroad Safety Culture Scales, there were a high number of comparisons. It was probable that some tests of statistical significance would be false-positives; that is, differences among groups would register as statistically significant when, in fact, the groups did not actually differ from each other. A multiple-stage process addressed this threat to correct interpretation:

1. First, an overall 2x3 MANOVA multivariate analysis was conducted with date (baseline and final) and site (1, 2, or 4) as independent variables and the scale averages for all the Railroad Safety Culture Scales (excluding the Safety Behavior Scales) as the dependent variables. This allowed for a single significance test increasing the confidence that observed differences were due to an effect rather than the high number of comparisons. If significant, this test determined if there were any changes to be investigated further. Then a 2x3 ANOVA was performed for each scale to see which scales had a significant interaction between date and site.
2. Given that step 1 revealed significant interaction, for each site a one-way ANOVA was performed with date (baseline and final) as the independent variable, and the scale averages for all the Railroad Safety Culture Scales (excluding the Safety Behavior Scales) as the dependent variables, and tenure and location as covariates. This determined if a given site had any significant changes over time. Then, for each site, a 2x1 t-test was performed for each scale to see which scales significantly increased, decreased, or did not change.
3. Finally, tests were performed to determine the 2x2 interaction for each pair of sites with date (baseline and final) and site as the independent variables and the scale averages for all the Railroad Safety Culture Scales (excluding the Safety Behavior Scales) as the dependent variables. This tested if the slopes over time were different between pairs of

sites. Then a 2x2 ANOVA was performed for each scale to see which scales had a significant interaction between date and site. This analysis revealed if one site was changing more than another over time on a given scale, or if both of them were improving by the same amount.

Steps 2 and 3 were repeated for the Safe Behavior Railroad Safety Culture scales. They had to be analyzed separately from the other Railroad Safety Culture scales because Site 1 did not have baseline values, and thus had to be excluded. Also Site 4 did not complete the dispatcher scale. Respondents were asked to select which behavior scale to answer based on their job location (i.e., yard vs. road), so they did not all complete every behavior scale. As a result, the multivariate analysis in step 1 would have resulted in list-wise deletion of all the respondents who only worked in one place. Instead, three separate 2x2 ANOVAs were used to avoid potential bias and loss of power. To control for family-wise error rate in the absence of a multivariate analysis, the Evaluation Team used the Bonferroni correction for the analyses of Safety Briefings, Road, and Yard. This set critical *p* in each ANOVA to 0.0167.

For the one demonstration site that completed its own survey, the Evaluation Team manually mapped similar scales between their survey and the C^{3}RS survey. Individual responses were not provided to the Evaluation Team, so statistical comparisons were not feasible.

For the survey questions on Impact and Sustainability, comparisons over time could be performed only for Sites 1 and 4, as Site 2 did not include those questions in their baseline. The three steps above were used to perform a 2x2 MANOVA, one-way ANOVAs for each site, and 2x2 interaction tests.

To analyze the C^{3}RS Reporting Questions, for each site, 2x2 chi-square tests were used for each question, comparing "yes" answers to "no" or "I don't know."

4.8 Corporate Archival Data

Each railroad's safety department provided archival corporate data.

- Site 1 provided human-factors incident data, including about 3 years of historical data before C^{3}RS began.
- Site 2 provided incident and injury data, including about 4 years of historical data before C^{3}RS began.
- Site 3 provided incident, injury, and discipline hearing data, including about 4 years of historical data before C^{3}RS began.
- Site 4 provided incident and injury data, including about 4 years of historical data before C^{3}RS began.

4.8.1 Incidents

Events that involve moving trains and damage to railroad equipment are called incidents. These events include collisions, run-through switches, and derailments, with the latter accounting for the large majority. A derailment is any incident in which at least one wheel of a railroad car, locomotive, or piece of equipment comes off the track. Some derailments can be catastrophic,

involving many injuries and huge amounts of equipment damage, but most derailments occur in the yard and involve minor repairs. FRA requires that railroads submit reports of all incidents costing the carrier greater than the reporting threshold (e.g., $9,200 in 2010). [29]

In this study, to obtain as much data as possible, both reportable and non-reportable incidents were included and analyzed. Even though only incidents below the FRA threshold could be reported to C^{3}RS (at some sites per their negotiated IMOU), corrective actions could have an impact on derailments both above and below the threshold. FRA also required railroads to collect data on incidents, investigate their causes, and record them in their corporate databases. As part of their investigation, railroads also had to classify incidents as being primarily either human factor or non-human factor caused. Human factor-caused incidents are those that have an employee human factor as the cause of the incident in the report. According to FRA, this includes causes such as: failure to properly secure the hand brake on car(s); employee asleep; flagging, improper, or failure to flag; radio communication, failure to comply; movement of engine(s) or car(s) without authority; failure to comply with restricted speed or its equivalent not in connection with a block or interlocking signal; and switch improperly lined. Non-human-factor incidents are due to other causes such as track geometry, signals, mechanical and electrical failures, and highway-user crossings.

When it began, C^{3}RS was hypothesized to reduce derailments:

- Site 1's C^{3}RS program focused on close call derailments and the actions humans take to cause them, so human-factors derailments were a candidate for possible reduction. Minor derailments (those under the FRA threshold) were allowed to be reported to C^{3}RS.
- At Site 2, the C^{3}RS outreach emphasized run-through switch close calls, because historically many derailments have been caused by switches that had been previously run-through and damaged but not reported to management. Site 2 provided a separate corporate data file including only the derailments they identified as being caused by run-through switches.
- At Site 3, derailments were not the primary focus of the C^{3}RS program; in fact, they could not report minor derailments to C^{3}RS, but they did focus on miscommunications and close calls in the yard, allowing some potential for reduction.
- At Site 4, they received some derailment-related C^{3}RS reports so they provided historical derailment data to the evaluation.

4.8.2 Discipline Hearing Data

When railroad employees violate a rule, they may be subject to a discipline hearing to determine if they are to be disciplined (e.g., suspension without pay or employment termination). Since C^{3}RS programs offered protection from discipline for certain rules, there is the potential for fewer discipline hearings. Hearings require substantial time and travel for managers, so fewer hearings implies reduced financial costs for the railroad.

Only Site 3 provided archival data on the number of discipline hearings per year for 4 years before C^{3}RS and 5.7 years after.

---

[29] 2011 FRA Guide for Preparing Accident/Incident Reports. Accessed online June 15, 2016.

It is reasonable to hypothesize that with immunity, workers may become more careless, and incidents and injuries could increase. The analysis of the other corporate data tested that hypothesis and also looked for increases.

4.8.3 Injuries

Railroad employees are sometimes injured during their job duties. While C^{3}RS has not had a large focus on injuries, some C^{3}RS corrective actions deal with unsafe conditions for employees and improved situational awareness. Therefore, there was some potential for C^{3}RS to reduce injury occurrences. Railroads are required by FRA to collect and record data on employee injuries and their impact on days off work. These data are organized by department. Since the railroads chose to only include transportation employees in the C^{3}RS demonstration pilots, the analysis focused on transportation employee injuries. These data were provided by Sites 2, 3, and 4. (Later in the C^{3}RS program, mechanical and engineering departments began to join, but they were not included during the demonstration period.)

The design of C^{3}RS was agnostic about the kinds of reports that employees could report. They could just as well have reported more events that led to personal injuries. Except for the exclusions described in the MOU and IMOUs, the reporting process was open to any kind of report that the employee wanted to submit related to safety. However, during the demonstration employees only submitted some reports related to injuries.

4.9 Occurrence Frequency Data Analysis Strategies

The Evaluation Team reviewed other Volpe Center evaluation studies (Zuschlag et al., 2012), and consulted with statistical experts, to determine the best way to analyze the frequencies of occurrence data (counts of derailments, discipline hearings, injuries, and C^{3}RS reports).

4.9.1 Selecting Impact Metrics

The Evaluation Team assessed C^{3}RS-analyzed-cases, corrective actions, and interview and field notes data to determine which corporate data might be analyzed for impacts of C^{3}RS. During phased interviews, labor and management respondents were asked what types of safety impacts could be expected given: 1) the types of risk at the site; 2) the categories of cases they had analyzed; and 3) the corrective actions that were implemented. The Evaluation Team then verified the extent to which the corrective actions mentioned by the respondents had been implemented. In addition, railroad experts from the sites and the C^{3}RS Implementation Team were consulted to further verify the likely impact of such mentioned corrective actions. After the verification activity, the Evaluation Team conducted statistical analyses to determine if significant trends in the rates could be detected. The archival corporate data were also explored to see if any trends or changes occurred at times other than at the start of C^{3}RS; then, those trends were discussed with the PRTs and management at the sites to determine the likely cause.

4.9.2 Occurrence Frequency Data Analysis

A given series of occurrences over time (e.g., derailments, discipline hearings, injuries, and C^{3}RS reporting rates) at a site may be adequately represented as a Poisson process, where occurrences randomly and independently accumulate at a nearly constant rate. Independence means that each occurrence is neither related to the previous nor the subsequent one (Nelson, 2003).

The correlations for adjacent gap times were low; this supports the independence assumption (Cohen, Cohen, West, and Aiken, 2002). "Gap times" are the amount of time between adjacent occurrences. The longer the gap times, the lower the occurrence rate. A lack of independence would be shown as long gaps tending to follow long gaps, and short gaps following short gaps. Independence means knowing the length of the current gap does not help you predict the length of a subsequent gap. That is, the correlation is near zero. To check this assumption the Evaluation Team calculated the autocorrelation for the natural log of the adjacent gap times. (For example, Site 1 adjacent derailments had a Pearson Correlation = 0.10, p=0.06, and for Site 3 adjacent derailments had a Pearson Correlation = -0.01, p=0.86.)

Consequently, one may expect occurrences are adequately modeled as a Poisson process, where the probability of an occurrence is constant within each gap between occurrences (Maguire, Pearson, and Wynn, 1952).

To assess the impact of C^{3}RS, the Evaluation Team compared rates of occurrences before and after key events in the program:

- Before and after the C^{3}RS program started.
- Before and after the C^{3}RS program implemented corrective actions.
- Before and after some date recommended by stakeholders, i.e., if interviewees said that they observed safety improving after a certain date.

In addition, Poisson Regression testing modeling was used at Site 2. The dates of the C^{3}RS time block coincidentally corresponded with several winters with unusually large amounts of snowfall. Site participants indicated that large snowfalls can cause extra derailments. To account for this potential confound, the Evaluation Team collected data on the inches of snowfall recorded by the National Oceanic and Atmospheric Administration.[30] The Poisson Regression test modeling allowed the Evaluation Team to use the inches of snowfall as a covariate when determining whether the reductions during C^{3}RS were significant.

The analyzing and interpreting data sought to determine whether findings were both statistically significant, i.e. convincing, and operationally substantial. Statistical significance is a measure of the likelihood that an observed difference might be nothing more than random fluctuation. However, additional understanding was needed because very small differences might show up as statistically significant depending on the measurements involved. Thus, what was also needed was an estimate of whether the observed difference was substantial enough to matter from a practical point of view with respect to people's lives and organizational processes (Davidson, 2005).

One potential situation that might arise was a continued improvement that started before C^{3}RS began and was not influenced by C^{3}RS. Rates for after a C^{3}RS key event would appear better than before, but it would be incorrect to attribute it to C^{3}RS. To address this possibility, cumulative incident plots were created to help the Evaluation Team identify when in time rate changes occurred and compare them to C^{3}RS's start time. Section 4.9.5 describes cumulative incident plots.

---

[30] National Oceanic and Atmospheric Administration (NOAA) Annual Climatological Summary accessed online at noaa.gov. The Evaluation Team used the NOAA data type: "total snowfall (TSNW)" per month.

4.9.3 Normalizing Data

To compare occurrence rates, the weight of time blocks needs to be proportional to the work completed (say, cars moved) instead of just the raw time. Normalization attempts to do this by dividing the event frequency by the work completed in a block. This is especially important for carriers that experienced large fluctuations in workload during the demonstration pilot. Following railroad practice, because most incidents occur in the yard, number of car-moves is generally used to normalize the frequency of incidents in order to account for the different volumes of traffic in different yards or in the same yard during different time periods. For example, incident rates can be calculated as the number of incidents per 100,000 car-moves at a location. Car-moves are the number of cars moved through a yard within a given length of time. Injuries data and discipline hearing data could be normalized using employee worker hours. The Evaluation Team worked with the sites to identify situations where normalization was needed. At Site 1, there were large fluctuations in work during the time periods being analyzed, so normalization was critical. At other sites, the quantity of work was consistent from year to year, so normalization was not needed, and the normalization data were not provided by the site. In those cases, rates were calculated as the number of occurrences per unit of calendar time.

For the C^3RS reporting data, the number of eligible employees was used to normalize the data. This was needed because: 1) one site had a large expansion of reporting eligibility in the middle of their demonstration, and 2) the Evaluation Team wanted to be able to consistently compare rates among sites whose C^3RS program size varied greatly.

4.9.4 Comparison Data

To help assess whether any observed changes were due to C^3RS and not due to other sources, such as another safety program or an extraneous variable (like snowfall), the C^3RS demonstration site data were compared to other locations with no or weak C^3RS implementation. Figure 13 shows the logic of such comparisons, such as demonstration site versus non-demonstration site. The analyses included comparing a C^3RS site with another location, such as another service unit or yard within the same railroad where the safety process was not implemented.

With the comparison data available, the chief statistical analysis was the relative performance of the treatment and comparison data and the presence of a statistically significant change over time at each location. If the treatment site's rate of occurrences changed significantly while the comparison site's rate did not, it suggests that the change was due to C^3RS. If the changes in safety were unique to the C^3RS site and not in the comparison data, then C^3RS was specifically impacting it. In contrast, if there were no differences in any observed change at treatment and comparison sites then there was no reason to believe that C^3RS had an impact (Rossi et al., 1999; Shadish, Cook, and Campbell, 2002).

Site 1 provided data from other geographic locations within the same region that were not involved in C^3RS. The site told the Evaluation Team which other locations had similar characteristics as the C^3RS demonstration pilot location revealing two possible locations to use as comparisons. Then the Evaluation Team asked the site if the other locations had any safety programs that might have impacted derailments. One potential comparison site did have another competing safety initiative occurring in parallel, so it was eliminated as a comparison site.

Site 4 provided comparison data from the entire region and sorted all the derailments as having occurred in the C^{3}RS demonstration covered areas vs. non-C^{3}RS areas. For Sites 2 and 3, comparison data were not available.

4.9.5 Cumulative Incidence Plots

A method for visually identifying potential changes in the rates of occurrences is through the use of cumulative incident plots (Nelson, 2003; Cook and Lawless, 2010). In such plots, the cumulative number of occurrences was plotted against the cumulative amount of exposure (e.g., time, worker hours, or cars moved). This provided a means to visually assess trends which might be difficult to perceive in the tabled data or monthly occurrence bar charts (histograms).

Cumulative number of incidence plots always have an upward trend, and the *steepness* or slope of the cumulative incidence plot is equal to the occurrence rate at a given point in time (or other cumulative exposure like cars moved). A generally straight line trend in the plot, as shown in Figure 18(a) indicates a constant occurrence rate, and by implication, no observed impact from a safety process. A cumulative incidence plot whose slope decreases indicates the incidents occurred less often, a positive safety change, as shown in Figure 18(b). Alternatively, an increasing slope indicates increasing rates, as in Figure 18 (c), i.e., the incidents occurred more often. An "elbow" in the plot indicates a step change in the incident rate. The location of the elbow on the x-axis shows when the change occurred (Zuschlag et al., 2012).

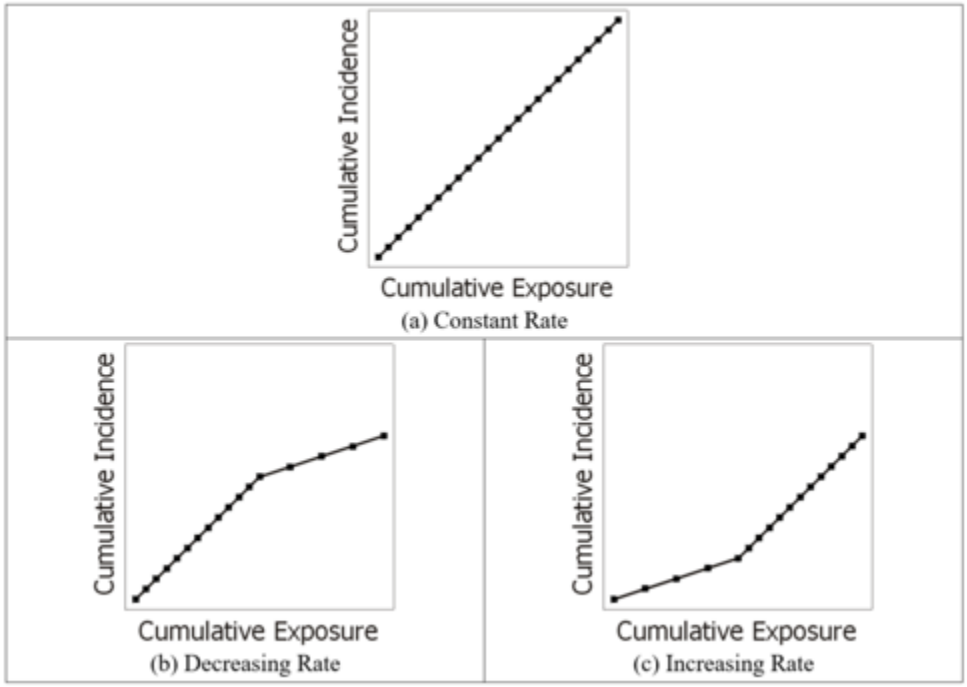


Figure 18. Interpretation of Cumulative Incidence Plots

5. Results

The demonstration pilot sites asked that their respective data be kept confidential; therefore, the labels for the sites were disguised (Sites 1, 2, 3, and 4), and the assignment of labels was not the sequence in which the sites joined the research project, which is public information. However, to avoid confusion, site designations are consistent throughout the report, so Site 1 always refers to the same demonstration pilot site; Site 2 always refers to the same site, and so forth. The assignment of numbers remains consistent with past evaluation reports.

This section of the report includes the results of the evaluation data analysis. The results in this section are presented in the same order as described in the Methods chapter and in Table 5. There are results related to: C^{3}RS reporting data; corrective action tracking data; interviews and other qualitative data (implementation interviews, phased interviews, project records, and field notes); the C^{3}RS railroad safety culture survey, and corporate archival data.

5.1 C^{3}RS Reporting Data

Figure 19 compares the average monthly C^{3}RS reporting rates among the four demonstration sites.[31] The Evaluation Team received the number of report-records collected per month per site. Those rates were normalized using the approximate number of employees eligible to report to C^{3}RS per site. For example, at Site 1 eligible reporters submitted approximately 1 C^{3}RS report every month per 100 employees, meaning if they had 500 eligible employees, they submitted about 5 reports a month. Comparing among sites, Site 4 showed the highest reporting frequency at approximately 3 reports per month per eligible employee (e.g., if Site 4 had 500 eligible employees, then the reporters submitted an average of 15 reports a month).

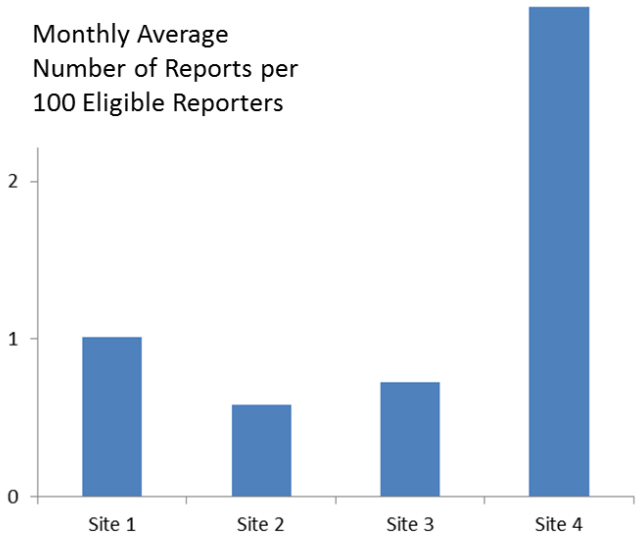


Figure 19. Average Monthly Reporting Rates

[31] For purposes of this report, the term "trends" refers to the actual close call reports submitted and accepted. It does not refer to the population of all close call events in the four sites that occurred during the evaluation period.

The Evaluation Team examined trends in the frequency of reporting from month to month for each demonstration site as an indication of successful implementation. Figure 20 shows the cumulative number of reports over time, for each site, again normalized by the number of eligible reporters. As explained in Section 4.9.5, cumulative charts were useful for showing increases and decreases in rates. A slope that became more vertical indicated that reports became more frequent, and a slope that became more horizontal showed that reporting became less frequent. Three of the four sites showed statistically significant changes in reporting rates, using F-tests for the ratio of two Poisson occurrence rates (Nelson, 2003). Table 13, Table 14, Table 15, and Table 16 show the following results. (For an explanation about how to read cumulative charts, see Section 4.9.5.) Note that Sites 2 and 3 only provided 3 years of data to the Evaluation Team while Sites 1 and 4 provided data for their whole demonstration. The reason for this data discrepancy was changes in record keeping by the C^{3}RS Third-Parties.

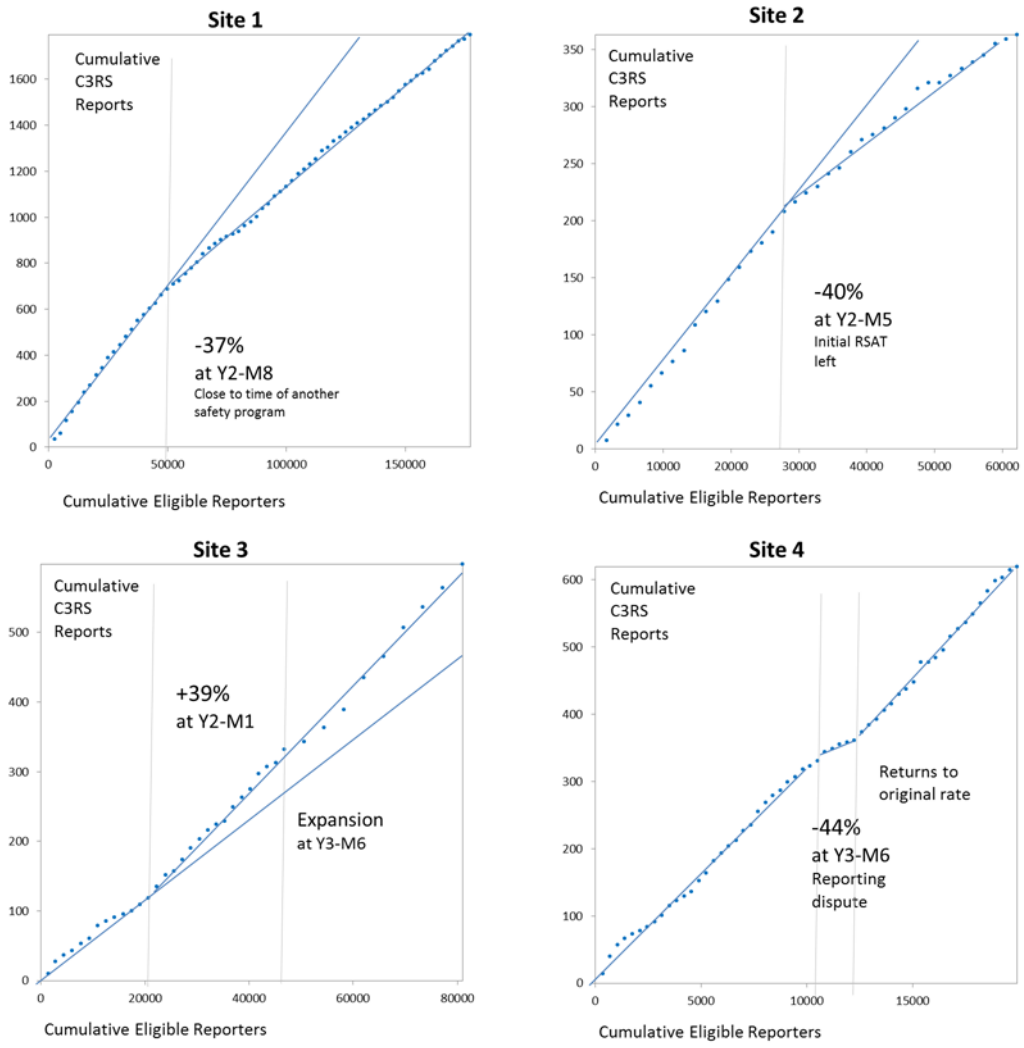


Figure 20. Cumulative Reporting Rates[32]

[32] In the figure Y=year and M=month, so "Y2-M8" refers to the eighth month of the second year of that site's C^{3}RS program.

Sites 1 and 2 experienced a slowing of reporting during their second year. Specifically, Site 1 had a 37% significant decrease in reporting rates ($F(2216,1374) = 1.58$, $p<0.001$) and Site 2's reporting rate significantly decreased 40% ($F(312,416) = 1.65$, $p<0.001$). The decrease in reporting for Site 1 occurred around the time another large safety program was initiated. The decrease at Site 2 coincided with personnel changes at the Third-Party, which may have caused eligible reporters to be hesitant to report. These extraneous events may have affected the observed decreases. The Evaluation Team looked at other reporting rate sources of data for Site 2, and found that higher reporting may have resumed later in the pilot.

In contrast, Site 3 had a 39% significant increase in reporting frequency at the beginning of the second year ($F(238,960) = 1.38$, $p<0.001$). This may have been related to the site needing time to educate eligible reporters and build their trust in C^{3}RS. However, when Site 3 expanded the number of eligible reporters able to report, the rate of reports per eligible employee remained constant.

Site 4's reporting remained mostly constant (non-significant 7% increase, $F(662,516) = 1.06$, $p=0.23$), except for a couple of months in the middle of the demonstration. This indicated that employees continued to report at a fairly constant rate and had no significant linear changes.

Table 13. Site 1 Reporting Rates

Phase	**Statistical Metric**	**C^{3}RS Site**
Early C^{3}RS (1.75 years)	Average number of C^{3}RS Reports per month per 100 eligible reporters	**1.37**
Later C^{3}RS (4.3 years)	Average number of C^{3}RS Reports per month per 100 eligible reporters	**0.87**
	Percent Change	**-37%**
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**<0.000**

Table 14. Site 2 Reporting Rates

Phase	Statistical Metric	C³RS Site
Early C³RS (1.4 years)	Average number of C³RS Reports per month per 100 eligible reporters	**0.75**
Later C³RS (1.8 years)	Average number of C³RS Reports per month per 100 eligible reporters	**0.45**
	Percent Change	**-40%**
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**<0.000**

Table 15. Site 3 Reporting Rates

Phase	Statistical Metric	C³RS Site
Early C³RS (1.1 years)	Average number of C³RS Reports per month per 100 eligible reporters	0.57
Later C³RS (2.1 years)	Average number of C³RS Reports per month per 100 eligible reporters	0.80
	Percent Change	**+39%**
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**0.001**

Table 16. Site 4 Reporting Rates

Phase	Statistical Metric	C³RS Site
C³RS (4.75 years)	Average number of C³RS Reports per month per 100 eligible reporters	**3.10**

These tables show that two of the sites had decreases in reporting about 38%, one had an increase of 39%, and one did not change. Reasons for the change at the three sites differed, as did the directions. However, when the change was observed, it showed a range of 40 percent.

The Third-Party also provided information on the categories of C³RS-analyzed-cases received by each site and compared them across sites to determine similarities and differences. One cannot assume that these are an exact representation of all the close call events that occurred; these are simply the categories that employees chose to report. This means that the highest frequency category at a site may not be the biggest safety problem at the site. The categories of reports submitted may have been influenced by the incentive of receiving disciplinary protection as outlined in their IMOU, the emphasis during C³RS training, and/or communications from the PRT.

Looking at Figure 21 across the four sites, excess speed was the most frequent category, followed by switching and derailments. Many corrective actions were implemented to address these safety concerns (for examples, see Section 5.2 below). Looking at Figure 22, one can see that for each site differences existed in the categories of close calls reported. Three of four sites reported excess speed, and all sites reported switching. Across the sites, the definitions of analyzed-case categories were not consistent, so the Evaluation Team consolidated them into the following definitions using the analyses provided in this report:

- **Excess Speed:** Cases where a crew exceeded a temporary or permanent speed restriction.
- **Switching:** Cases involving switching, such as a run-through or misaligned switch.
- **Derailment:** Cases where a crew experienced a minor derailment, below the FRA reporting threshold.
- **Collision:** Cases where there was a near miss with another vehicle or person.
- **Proper Authority:** Cases where the train did not have permission to operate.
- **Dispatcher:** Cases involving dispatcher close calls or errors.
- **Others:** All other categories of cases.

In summary, while excess speed, switching, and derailment were the most frequently reported across the demonstrations, there was variability between the sites in the most frequently reported categories.

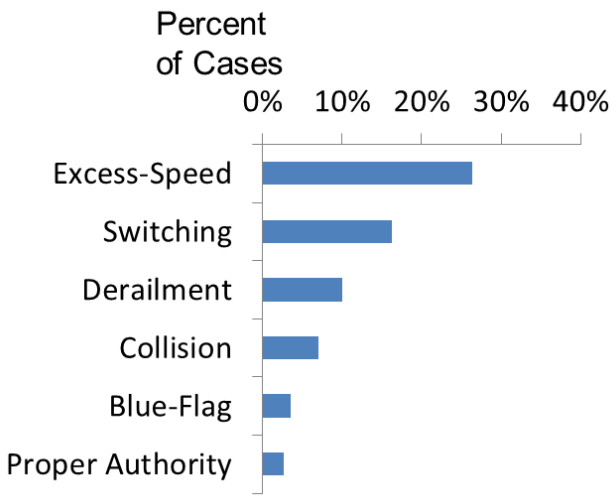


Figure 21. Overall Percentages of Categories of C³RS-Analyzed-Cases Across Four Sites

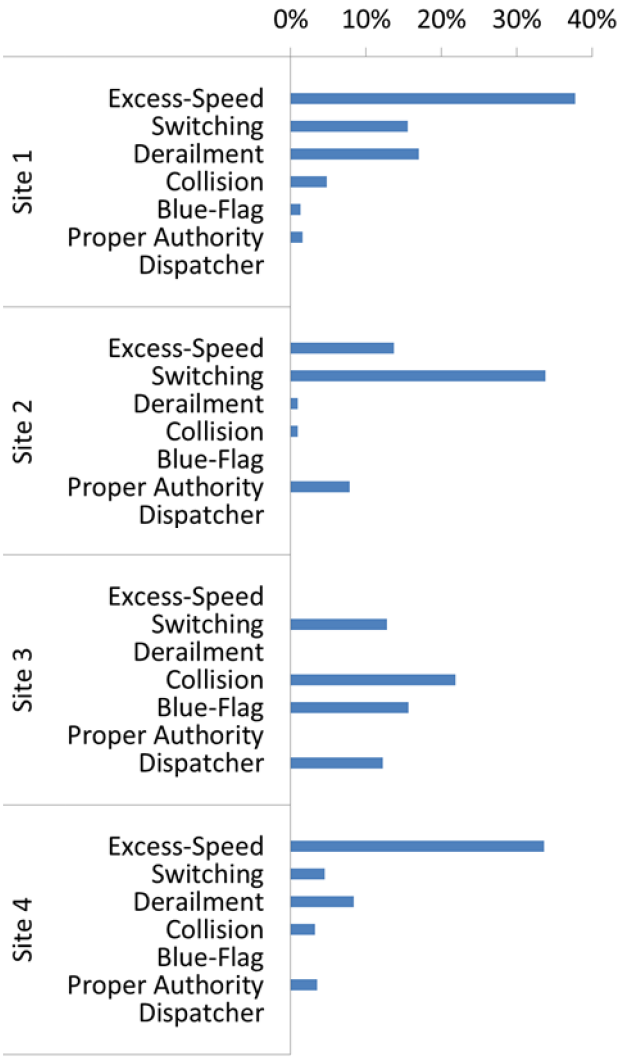


Figure 22. Percentages of Categories of C³RS-Analyzed-Cases Within Sites

Figure 23 shows the proportion of C^{3}RS-analyzed-cases that were "reporter-known" events for Sites 1, 2, and 4. (This data was not available for Site 3.) Reporter-known events are minor incidents where the reporter submits a report to the C^{3}RS Third-Party *as well as* informs the railroad directly, i.e. the railroad knows the identity of the reporter. For example, if the reporter provides information directly to the railroad about a run-through switch, then the railroad can fix the switch and prevent a derailment. Reporter-known events tend to include minor violations, where discipline can be avoided when a report is submitted, so it is safe to report on such situations as run-through switches and minor derailments (with damage below the mandatory FRA reporting threshold). "Reporter-unknown" events are those events where the reporter chooses to submit only to the Third-Party. Reporter-unknown events include situations such as excess speed and near collisions.

The total number of submitted reports across three railroads consisted of 44% (720 reports), where the identity of the reporter was known to the railroad and 56% (910 reports) where the identity of the reporter was not known. This means that 56% of the close call events reported would not have been known by the railroad without a C^{3}RS report, which shows that the railroads are gaining new information.

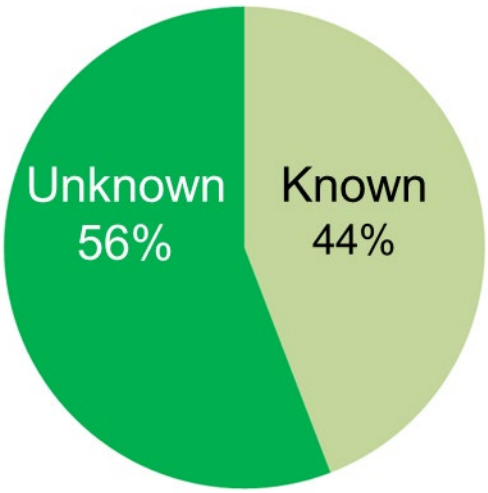


Figure 23. Close Calls vs. Known Events (Sites 1, 2, and 4)

5.2 Corrective Action Tracking Data

Qualitative corrective action data showed the extent to which the program and corrective actions were implemented. Content analysis of the corrective actions revealed their mapping to the top analyzed-case categories: excess speed, switching, derailments, collisions, and blue flags. Examples of corrective actions in each of these categories are shown in Table 17–Table 20. The tables include four columns. The first column contains the analyzed-case category assigned by the Third-Party. The second column is a short description of the PRT's analysis of the case. The third column contains the corrective action implemented by the site. The fourth column contains the impact goal the site was trying to achieve by implementing this corrective action. Appendix H. Example Corrective Actions contains a longer list of examples with explanations of the cases and corrective actions.

Sites had a challenge as corrective action data were not consistently well organized. The sites had corrective action records in multiple places, and not all local actions were written down. The records did not always clearly identify who was responsible to implement the action, whether or not the action was fully implemented, and when it was implemented. This put all the sites at risk of losing valuable information as participants turned over. Because of the gaps in recordkeeping the Evaluation Team used interview data to fill out the details. To verify corrective action data reliability, the results were shared with the demonstration sites. Despite the challenges with record-keeping, all the sites implemented numerous corrective actions. Their actions were concentrated in the categories where they had the most reports.

Table 17. Example Corrective Actions for Site 1

Category	C^{3}RS-Analyzed-Case	Corrective Action	Possible Impact
Excess Speed	Difficult to read track warrant paperwork	Reformatted track bulletin	Reduced speeding Reduced decertifications
Excess Speed	Many speed restrictions in short distance	Consolidation of slow orders	Reduced speeding Reduced decertifications
Derailment	Foreman throwing switch and not seeing train was still in the area	Added a camera to that part of the yard	Reduced derailments
Derailment	Forgetting to remove derails after completing of maintenance	New policy to assure derails are removed	Reduced derailments
Derailment	Unable to tell which way switch is lined	Reflective switch targets	Reduced run-through switches Reduced derailments
Run-through Switch	Run-through switches in yard	Added content to training	Reduced run-through switches Reduced derailments

Table 18. Example Corrective Actions for Site 2

Category	Analyzed-Case	Corrective Action	Possible Impact
Excess Speed	Difficult to see track warrant paperwork	Added clip to cab to keep paperwork in sightline	Reduced speeding Reduced decertifications
Excess Speed	Forgetting about speed restriction after station stop	Reminder signs	Reduced speeding Reduced decertifications
Derailment	Miscommunication in yard during reverse moves	Improved communication protocol	Reduced derailments
Run-through Switch	Unable to see which way switch is aligned	Improved lighting	Reduced run-through switches Reduced derailments

Category	Analyzed-Case	Corrective Action	Possible Impact
Run-through Switch	Unable to see which way switch is aligned	Painting switch	Reduced run-through switches Reduced derailments
Other	Crews unaware of important safety information when starting a new shift	Improvements and extensive education about job safety briefing	Reduced injuries

Table 19. Example Corrective Actions for Site 3

Category	Analyzed-Case	Corrective Action	Possible Impact
Excess Speed	Dispatcher error in labeling speed restrictions	Change dispatcher assignments so they are not changing between yard and road work in a single shift	Reduced speeding Reduced decertifications
Excess Speed	Difficult to read track warrant paperwork	Reformatted track bulletin	Reduced speeding Reduced decertifications
Derailment	Crews moving on protected track and hitting a derail	Performed training about rules for switch position and shoving	Reduced derailments
Derailment	Crews working in an unfamiliar yard	New policy that if have not worked in a yard in six months, re-training is required	Reduced run-through switches Reduced derailments
Run-through Switch	Miscommunication between yard control and yard crews	Improved radio procedures	Reduced run-through switches Reduced derailments
Collision	Vehicles fouling track in yard	Contractor requirement to not use cell phones in the yard	Reduced collisions Reduced injuries
Collision	Employees crossing track when walking from parking lot	Put up barriers	Reduced collisions Reduced injuries
Collision	Dump truck fouling track	Moved the dumpster	Reduced collisions Reduced injuries

Table 20. Example Corrective Actions for Site 4

Category	Analyzed-Case	Corrective Action	Possible Impact
Excess Speed	Many speed restrictions in short distance	Combined some permanent and temporary speed restrictions	Reduced speeding Reduced decertifications

Category	Analyzed-Case	Corrective Action	Possible Impact
Excess Speed	Crew leaving old paperwork in cab	Poster to remind crews to remove their paperwork at end of shift	Reduced speeding Reduced decertifications
Excess Speed	Confusing speed restrictions	Information meeting to review restrictions	Reduced speeding Reduced decertifications
Run-through Switch	Crews struggling with remote control switches	Flyer explaining how to activate remote control switch and identifying locations	Reduced run-through switches Reduced derailments

Different sites used some similar corrective actions for the top-three analyzed-case categories (Table 21). For example, three sites reformatted their track bulletins for crews to read and follow more easily, reducing the likelihood of excess speed.

Table 21. Correction Actions Similarity Across Sites

Category	Action	Site	Action	Site	Action	Site
Excess Speed	Reformatting track bulletins	1 2 3	Reminder tags, Modified signs	1 2 4	Consolidating slow orders	1 4
Switching	Improved communication	2 3	Improved switch visibility	1 2	Training	1 2 3
Derailment	Improved communication in yard (e.g., squawk box)	2 3	Visual checks (e.g., camera, checking with yardmaster)	1 3	Training	1 2 3

5.3 Interviews and Other Qualitative Data

This section discusses the results of the analysis of interviews and other qualitative data which included implementation interviews, phased interviews, project records, and field notes.

In Section 5.3.1, the ratings of the demonstration site's implementation, impact, and sustainability are presented. This includes ratings for the implementation fishbone diagram in Figure 16 and the summary assessments of impact and sustainability (as described in Section 4.6.2.4).

In Sections 5.3.1.5–5.3.1.6, after the sites are discussed separately, a discussion of the similar and different ratings on demonstration sites' implementation main factors is presented.

In Section 5.3.1.7, also after the sites are discussed separately, then a discussion of similarities and differences in sustainability across the demonstration sites is presented.

In Section 5.3.2, the industry-level sustainability fishbone diagram and ratings are presented. This includes ratings for the industry sustainability fishbone in Figure 17. The data used for this focuses on industry factors, not demonstration site factors (as described in Section 4.6.2.5).

5.3.1 Demonstration Sites' Implementation, Impact, and Sustainability Rating Results

In this section, each site's fishbone diagram and ratings, compiled from the qualitative data, are discussed. (Refer to Table 7.) Quantitative impact results from Sections 5.4 and 5.5 are included to show how the impacts and sustainability were related to implementation for each demonstration. Then, the four demonstrations are compared.

5.3.1.1 Site 1

Site 1 completed the activities outlined in Figure 3 as planned. Their employees observed close calls and submitted reports to C^3RS. The Third-Party processed their reports. Their PRT analyzed cases. The railroad reviewed and implemented some corrective actions. The railroad shared some information with the PRT and employees about what was implemented.

The ratings used to create the fishbone diagram for Site 1's implementation, impact, and sustainability are described below.

Implementation – For each main implementation factor, the Evaluation Team rated each detailed factor within it and counted the number of detailed factors with each rating. The counts are below (Table 22). For example, under FRA Responsibility there are five detailed factors; the team rated one detailed factor "fair," three detailed factors "good," and one detailed factor "very good." (Explanations of ratings the Evaluation Team assigned to every detailed factor are in Appendix F. Detailed Implementation Assessment Rubric for Each Site.) In addition, the right two columns, similar to the ratings shown in Table 7, show the computed average rating for each main factor. For each main factor, the qualitative average rating of very good, good, fair, or poor was determined by rounding the quantitative average to the nearest whole number. Using the same example of FRA Responsibility, the average rating for detailed factors under FRA Responsibility was "3" or "good." (This was calculated: ((2+3+3+3+4)/5=3.0).

In reviewing Table 22, one can see that overall the ratings were favorable in that 5 of the 8 main implementation factors received a rating of very good = 4 or good = 3. Only 3 of the main factor ratings were unfavorable, that is, fair = 2 or poor = 1. Of the five main factors that were under the demonstration site's control, they received two unfavorable ratings on shared responsibility and ability to implement corrective actions because of issues with communicating to the workforce about corrective actions, improving process efficiency, and tracking all the corrective actions. The factor "shared responsibility" had one "poor" rating, due to the inability of all internal and external stakeholders to renegotiate the IMOU and waivers after the demonstration was over. However, they received favorable ratings on cooperating on implementing corrective actions.

Table 22. Implementation Ratings at Site 1

Main Factor (number of detailed factors in parentheses)	Count of Detailed Factors with this Rating				Average Rating (rounded)	
	Poor = 1	Fair = 2	Good = 3	Very Good = 4	Quantitative	Qualitative
FRA Responsibility (5)		1	3	1	*3*	Good
Third-Party Responsibility (4)	1	1	2		*2*	Fair
Carrier Responsibility (5)		1	4		*3*	Good
Labor Responsibility (4)		1	2	1	*3*	Good
Shared Responsibility (8)	1	5	1	1	*2*	Fair
Ability to Implement Corrective Actions (3)		2	1		*2*	Fair
Effective Dispute Resolution (2)			1	1	*4*	Very Good
Perceived Value (4)			4		*3*	Good

Impact – Observed impact was rated as "substantial impact = 4" because Site 1 had a significant reduction in derailments (Section 5.5.1) and had improvements in safety culture (Section 5.4). Interviewees believed that C^{3}RS had a positive influence on safety and safety culture. The overall value for impact is placed at the head of the fishbone on Figure 25. Interviewees believed that C^{3}RS had a positive influence on safety and safety culture, thus a rating of "good" perceived value was also given in the implementation analysis above.

Sustainability – Sustainability was rated as "withdrew =1" because Site 1 left C^{3}RS a few years after the demonstration ended. At the end of their demonstration, Site 1 planned to continue. However, in later years, Site 1 had concerns about the Third-Party that FRA decided to use after the demonstration and FRA's plans for an anonymized shared industry close call database. Site 1's waiver expired, and they left FRA's C^{3}RS program unable to negotiate a revised IMOU with internal and external stakeholders. Also, in later years, as they considered re-joining, the management was concerned about including "known events" as part of their IMOU. In summary, they did not continue with C^{3}RS because the stakeholders could not resolve these conflicts. However, Site 1 did appreciate the value of close call reporting, and interviewees said that Site 1 later used similar concepts in their internal safety programs. (An assessment of their internal programs was out of the scope of this evaluation.) The overall value for sustainability is placed at the head of the fishbone on Figure 25.

Combining the results above (using the legend in Figure 24), the summary fishbone for Site 1 is in Figure 25. This figure contains the detailed implementation factors and a single separate value for both impact and sustainability. The value for impact is 4, and the value for sustainability is 1. This figure uses abbreviated labels for the main factors. This figure will be used later when

comparing across sites. (Due to additional information obtained in the post-pilot interviews, some detailed factors received a lower rating. The * symbol indicates that the rating displayed was lowered after the final evaluation because of this post-pilot information.)

Figure 24. Legend for Fishbone Rating Symbols

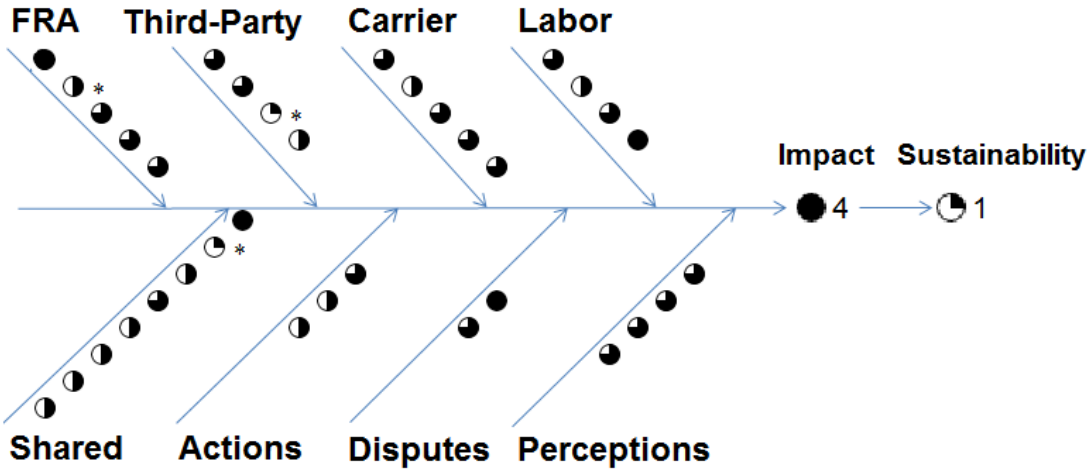


Figure 25. Site 1 Fishbone Combining Implementation, Impact, and Sustainability

5.3.1.2 Site 2

Site 2 completed the activities outlined in Figure 3 as planned. Their employees observed close calls and submitted reports to C^{3}RS. The Third-Party processed their reports. Their PRT analyzed cases. The railroad reviewed and implemented some corrective actions. The railroad shared some information with the PRT and employees about what was implemented.

See below for the ratings used to create the fishbone diagram for Site 2's implementation, impact, and sustainability.

Implementation – The layout of the Table 23 is similar to that of Site 1. The table contains counts of the number of Evaluation Team ratings for detailed implementation factors and the average rating for each main factor. (Explanations for the ratings of the detailed factors are in Appendix F. Detailed Implementation Assessment Rubric for Each Site.)

In reviewing Table 23, one can see that overall the ratings were favorable in that 6 of the 8 main implementation factors received a rating of very good = 4 or good = 3. Only 2 of the main factor ratings were unfavorable, that is, fair = 2 or poor = 1. Site 2 shared responsibilities were rated "good" on average. The "fair" average rating for labor responsibility was a result of periods of low reporting and a period of low PRT activity after a large PRT turnover. The "fair" rating for dispute resolution was a result of dispute after the demonstration period about whether a reporter

should be disciplined for multiple close calls in one day, during which one union considered withdrawing from C^{3}RS.

Table 23. Implementation Ratings at Site 2

Main Factor (number of detailed factors in parentheses)	Count of Detailed Factors with this Rating				Average Rating (rounded)	
	Poor = 1	Fair = 2	Good = 3	Very Good = 4	Quantitative	Qualitative
FRA Responsibility (5)		1	3	1	*3*	Good
Third-Party Responsibility (4)		2	1	1	*3*	Good
Carrier Responsibility (5)		1	4		*3*	Good
Labor Responsibility (4)		2	2		*2*	Fair
Shared Responsibility (8)		3	4	1	*3*	Good
Ability to Implement Corrective Actions (3)			3		*3*	Good
Effective Dispute Resolution (2)		1	1		*2*	Fair
Perceived Value (4)			4		*3*	Good

Impact – Rated as "substantial impact =4" because Site 2 had a significant reduction in derailments, although only when C^{3}RS was in high usage (Section 5.5.2). The site also observed improvements in safety culture (Section 5.4). Interviewees believed that C^{3}RS had a positive influence on safety and safety culture, thus a rating of "good" perceived value was also given in the implementation analysis above.

Sustainability – Rated as "continued =4" because Site 2 remained in the C^{3}RS program after their demonstration and was still involved as of the writing of this report. At the end of their demonstration, they needed to revise their IMOU to continue with C^{3}RS, and all the internal and external stakeholders agreed to it.

Combining the results above, the summary fishbone for Site 2 is in Figure 26. This figure contains the detailed implementation factors and a single summary value each for both impact and sustainability. This figure uses abbreviated labels for the main factors to enhance the visuals. The Evaluation Team will use this summary figure later when comparing across sites. (The legend for the symbols is provided in Figure 24.)

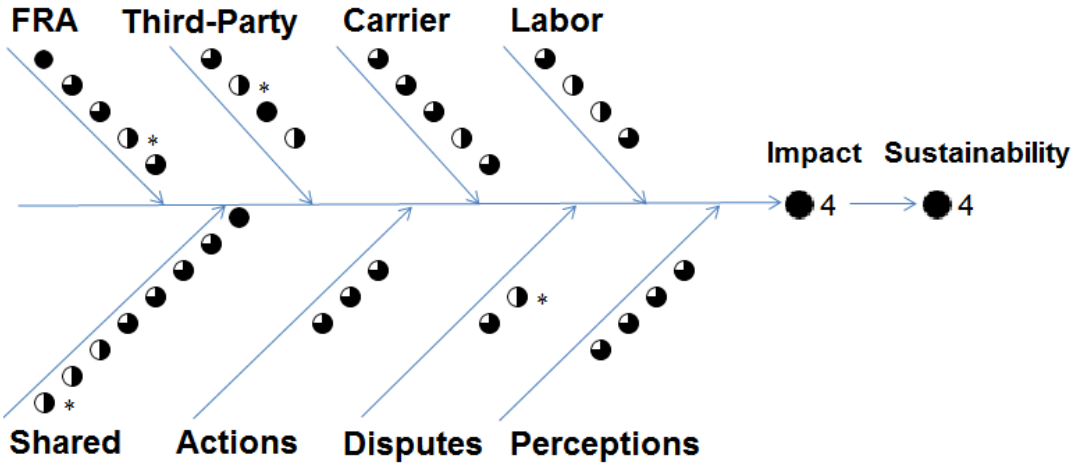


Figure 26. Site 2 Fishbone Combining Implementation, Impact, and Sustainability

5.3.1.3 Site 3

Site 3 completed the activities outlined in Figure 3 as planned. Their employees observed close calls and submitted reports to C^{3}RS. The Third-Party processed their reports. Their PRT analyzed cases. The railroad reviewed and implemented some corrective actions. The railroad shared some information with the PRT and employees about what was implemented.

See below for the ratings used to create the fishbone diagram for Site 3's implementation, impact, and sustainability.

Implementation – The layout of Table 24 is similar to that of Site 1. The table contains counts of number of Evaluation Team ratings for detailed implementation factors and the average rating for each main factor. (Explanations for the ratings of the detailed factors are in Appendix F. Detailed Implementation Assessment Rubric for Each Site.)

In reviewing Table 24 one can see that overall the ratings were favorable in that 7 of the 8 main implementation factors received a rating of very good = 4 or good = 3. Only 1 of the main factor ratings were unfavorable, that is, fair = 2 or poor = 1. Site 3 shared responsibilities were rated "good" on average. The "fair" average rating for labor responsibility was because the PRT wished the reporters would provide more information in their reports. It was also difficult for labor to promote the program to all geographic areas.

Table 24. Implementation Ratings at Site 3

Main Factor (number of detailed factors in parentheses)	Count of Detailed Factors with this Rating				Average Rating (rounded)	
	Poor = 1	Fair = 2	Good = 3	Very Good = 4	Quantitative	Qualitative
FRA Responsibility (5)			4	1	*3*	Good
Third-Party Responsibility (4)		2		2	*3*	Good
Carrier Responsibility (5)		1	3	1	*3*	Good
Labor Responsibility (4)		2	2		*2*	Fair
Shared Responsibility (8)		3	3	2	*3*	Good
Ability to Implement Corrective Actions (3)		1	2		*3*	Good
Effective Dispute Resolution (2)				2	*4*	Very Good
Perceived Value (4)			4		*3*	Good

Impact – Rated as "substantial impact=4" because Site 3 had a non-significant reduction in derailments that was operationally substantial (Section 5.5.3) and had improvements in safety culture (Section 5.4). Interviewees believed that C^{3}RS had a positive influence on safety and safety culture, thus a rating of "good" perceived value was also give in the implementation analysis above.

Sustainability – Rated as "continued=4" because Site 3 remained in the C^{3}RS program after their demonstration and was still involved as of the writing of this report. They renegotiated their IMOU, finding agreement with internal and external stakeholders, and expanded their program.

Combining the results above, the summary fishbone for Site 3 is in Figure 27. This figure contains the detailed implementation factors and a single summary value each for both impact and sustainability. This figure uses abbreviated labels for the main factors to enhance the visuals. This summary figure will be used later when comparing across sites. (The legend for the symbols is in Figure 24.)

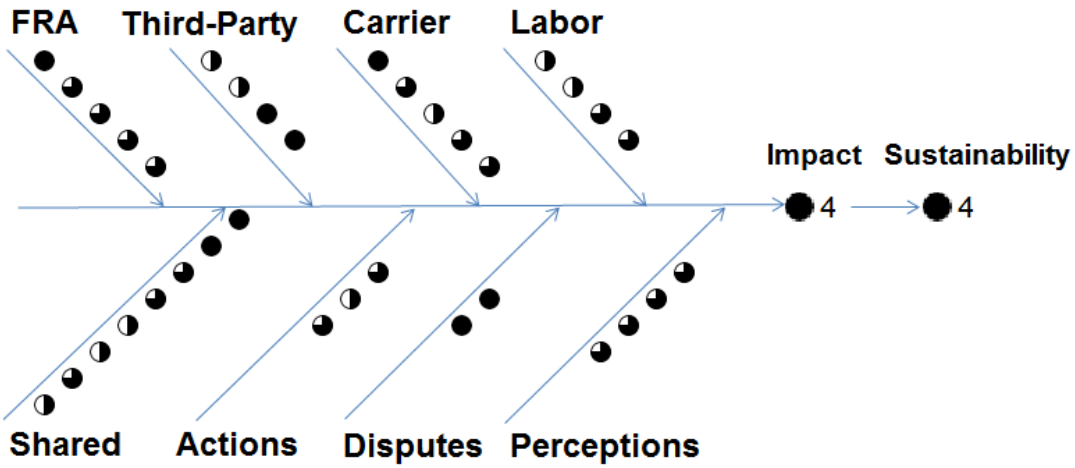


Figure 27. Site 3 Fishbone Combining Implementation, Impact, and Sustainability

5.3.1.4 Site 4

Site 4 completed the activities outlined on Figure 3 as planned. Their employees observed close calls and submitted reports to C^{3}RS. The Third-Party processed their reports. Their PRT analyzed cases. The railroad reviewed and implemented some corrective actions. The railroad shared some information with the PRT and employees about what was implemented.

See below for the ratings used to create the fishbone diagram for Site 4 implementation, impact, and sustainability.

Implementation – The layout of the Table 25 is similar to that of Site 1. The table contains counts of number of Evaluation Team ratings for detailed implementation factors and the average rating for each main factor. (Explanations for the ratings of the detailed factors are in Appendix F. Detailed Implementation Assessment Rubric for Each Site.)

In reviewing Table 25, one can see that overall the ratings were not favorable in that only 2 of the 8 main implementation factors received a rating of very good = 4 or good = 3. Thus, 6 of the 8 main factor ratings were unfavorable; that is, fair = 2 or poor = 1. Site 3 shared responsibilities within the site were all rated unfavorably. Management support of C^{3}RS varied, and the original Support Team did not fulfill its responsibilities. Some labor members put too much emphasis on "get out of jail free" and did not provide sufficient detail in their reports. Labor PRT members did not help management implement corrective actions. A few disputes arose in which management disciplined reporters. As a result, there were insufficient corrective actions implemented and little perceived value. The site also attempted to renegotiate their IMOU to resolve some of their implementation issues, but the negotiation between internal and external stakeholders was not successful.

Table 25. Implementation Ratings at Site 4

Main Factor	Count of Detailed Factors with this Rating				Average Rating (rounded)	
	Poor = 1	Fair = 2	Good = 3	Very Good = 4	Quantitative	Qualitative
FRA Responsibility (5)	1	1	2	1	*3*	Good
Third-Party Responsibility (4)		1	2	1	*3*	Good
Carrier Responsibility (5)	2		3		*2*	Fair
Labor Responsibility (4)	3	1			*1*	Poor
Shared Responsibility (8)	1	4	2	1	*2*	Fair
Ability to Implement Corrective Actions (3)	3				*1*	Poor
Effective Dispute Resolution (2)	2				*1*	Poor
Perceived Value (4)		4			*2*	Fair

Impact – Rated as "little impact=1" because Site 4 had no improvements in safety metrics (Section 5.5.4) and some improvements in safety culture (Section 5.4). Interviewees did not believe that C^{3}RS had a positive influence on safety and safety culture.

Sustainability – Rated as "withdrew=4" because Site 4 left C^{3}RS after the demonstration ended. They attempted to negotiate their IMOU, as mentioned above in relation to the implementation detailed factor, IMOU re-negotiation, but they did not reach a consensus between internal (labor and management) and external stakeholders (FRA and national labor).

Combining the results above, the summary fishbone for Site 4 is in Figure 28. This figure contains the detailed implementation factors and a single summary value each for both impact and sustainability. This figure uses abbreviated labels for the main factors to enhance the visuals. This summary figure will be used later when comparing among sites. (The legend for the symbols is in Figure 24.)

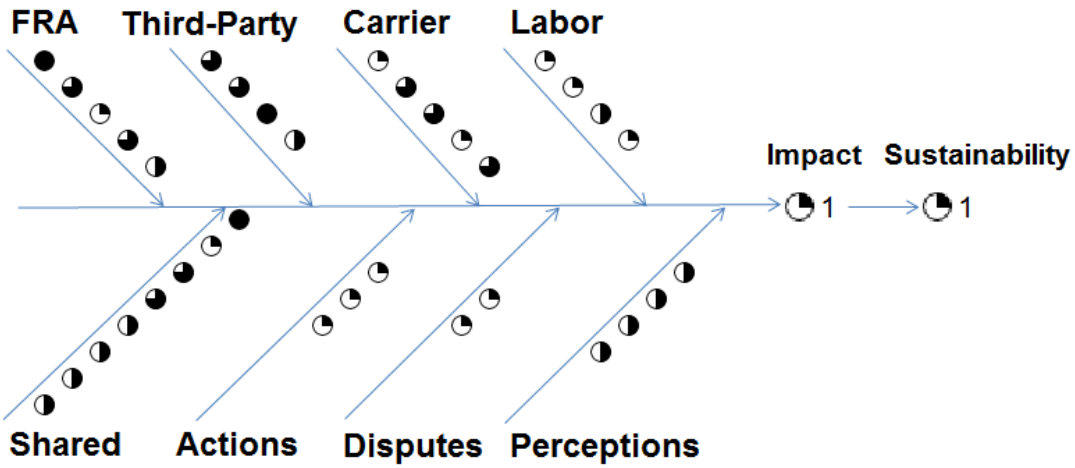


Figure 28. Site 4 Fishbone Combining Implementation, Impact, and Sustainability

5.3.1.5 Cross Site Implementation, Impact, and Sustainability Ratings

To facilitate cross site comparisons, Figure 29 reproduces all four sites fishbone diagrams.

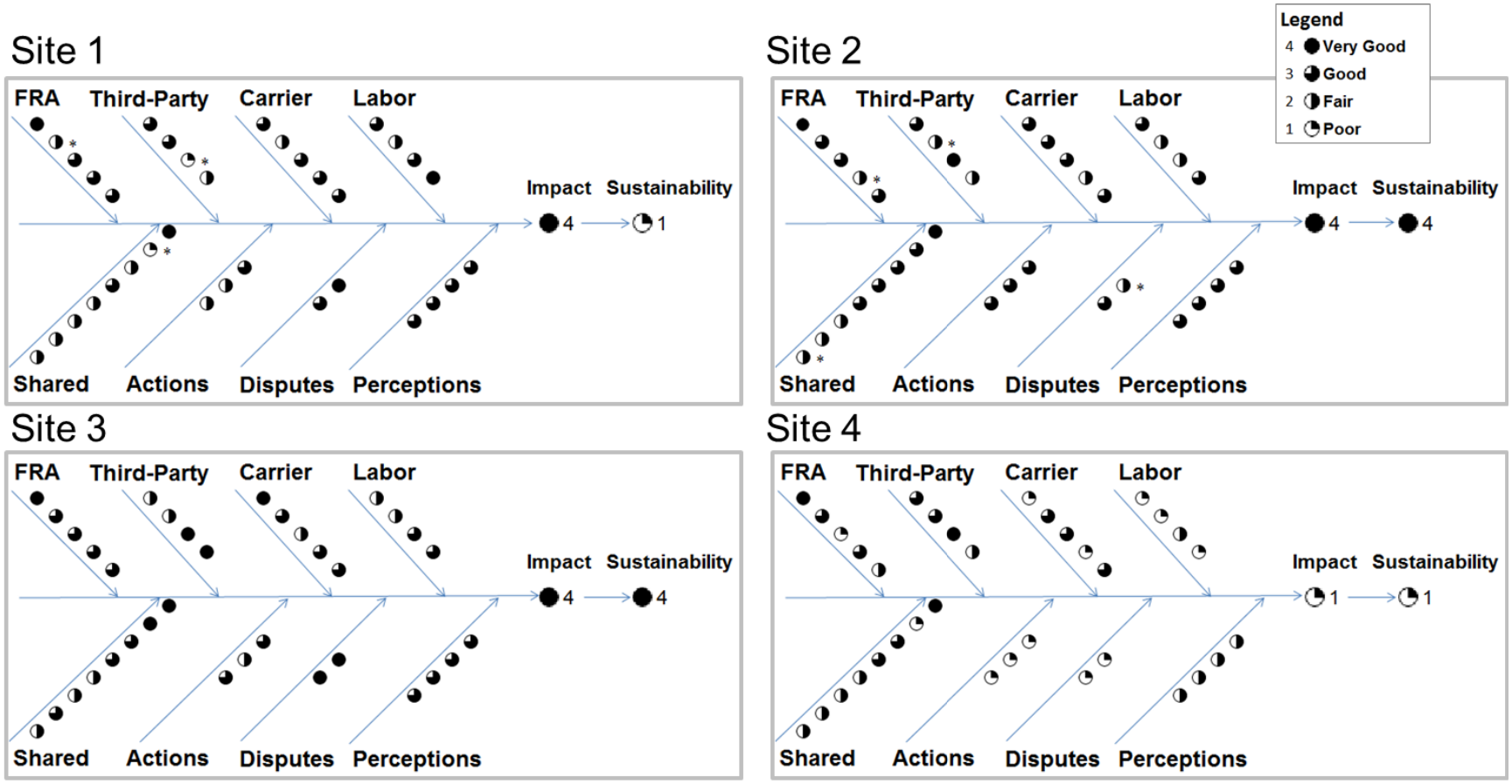


Figure 29. Summary Fishbone Diagrams Displaying Implementation, Impact, and Sustainability Across Demonstration Sites

Looking at the relationship between implementation, impact, and sustainability for the four sites in Figure 29, some observations can be made.[33]

- Sites 1-3 had many "good" and "very good" rated detailed implementation factors as well as a rating of "substantial impact." In contrast, Site 4 had less favorable detailed implementation factors and a rating of "poor" for perceived value and a rating of "little impact."
- Demonstration sites had different challenges. Sites 1-3 varied in which detailed factors were rated "good" and "fair."
- Sites 1-3 had some "fair" factors and still were able to implement corrective actions and have a "substantial impact."
- Sites 1 and 4, with a sustainability rating of "withdrew," both had some "poor" detailed implementation factors. Sites 2 and 3, with a sustainability rating of "continued" had no "poor" implementation factors.

[33] Full details on how each factor was rated are available in the appendices: Appendix F. Detailed Implementation Assessment Rubric for Each Site and Table 7.

5.3.1.6 Cross Site Common Challenging and Supportive Implementation Factors

Looking at the same detailed ratings data, but this time organized by factor, showed which implementation factors were challenging and which were successful across sites (Figure 30).

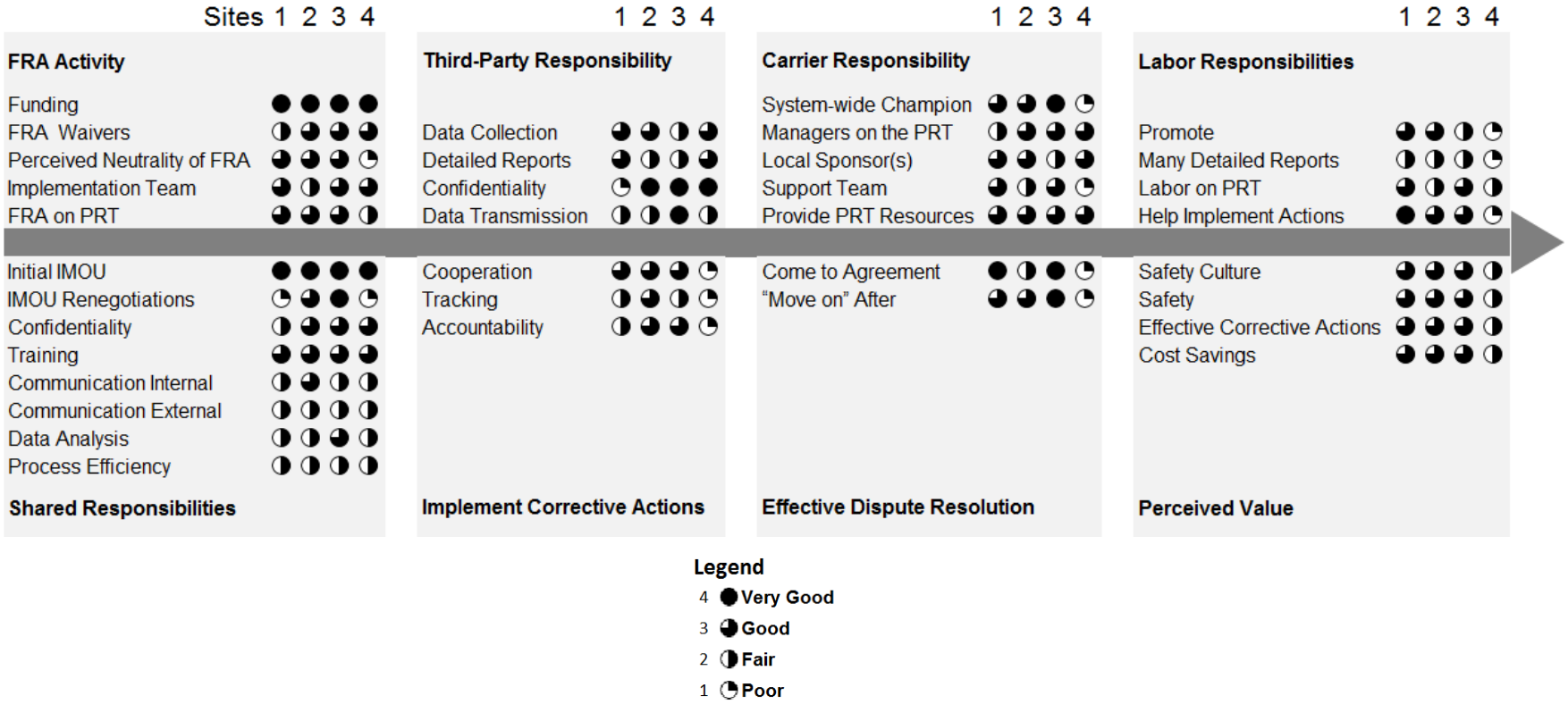


Figure 30. Detailed Implementation Ratings – Organized by Factors

FRA Activity received favorable ratings on average at all sites from the Evaluation Team. FRA provided consistent funding and helped establish new sites. In particular, FRA set up a Human Performance Division to provide assistance. However, three sites had a different detailed factor that was "fair" or "poor." FRA did not extend Site 1's waiver, due to disagreements on the Third-Party. Management at Site 4 felt that FRA was biased and favored labor during their disputes over discipline. Site 2 wanted more training for new PRT members and more reports from FRA about program accomplishments.

Third-Party Responsibility had a mix of ratings at each site from the Evaluation Team. It was Site 4's best main factor, because the Third-Party did a good job supporting Site 4 and did not cause Site 4's lack of impact and sustainability. For Sites 2 and 3, the Third-Party had to cut details for the sake of confidentiality, and the PRTs wished there was more detail. There were concerns about the efficiency of data transmission at three sites, which seemed to improve over time. After the demonstration, Site 1 cited concerns that the Third-Party would not maintain confidentiality for the carrier and individuals.

Carrier Responsibility received a number of "good" ratings across sites, showing carrier commitment to the program. In general, at Sites 1-3 management involvement was positive. They all had support from the senior levels of the carrier. Each of those three sites had "good" support from either the managers on the PRT or local management sponsors. Each site had detailed factors that received "fair" or "poor" ratings that were different from each other. Site 4 had two "poor" areas concerning the initial Support Team and the lack of a system wide champion.

Labor Responsibility received a mix of ratings across sites. For Sites 1, 2, and 3, labor on the PRT helped implement corrective actions. Site 4 had all "poor" ratings except for a "fair" job by the PRT. The detailed factor that all the demonstrations had unfavorable ratings concerned encouraging reporters to provide detailed reports. In general, interviewees said that this detailed factor improved over time as trust grew (except for at Site 4). However, even with some superficial reports, Sites 1, 2, and 3 developed and implemented corrective actions and improved safety impacts.

Shared Responsibility presented many challenges for all sites. The first two detailed factors concerned the IMOU, which involved stakeholders both within and external to the carriers. Agreement on the IMOU turned out to be one of the most challenging issues as it required cooperation with not only carrier labor and management but also with the external stakeholders including FRA, Third Party, and national labor organizations. This broad involvement was needed because C^3RS is FRA's national program and they strive for a certain level of consistency, therefore the IMOU is a document that requires all the internal and external stakeholders to agree on the way the program operates. While the "Initial IMOU" detailed factor tended to receive favorable ratings ("very good" for all sites) across the sites, the "IMOU Renegotiations" detailed factor did not. Sites 1 and 4 desired and were not able to renegotiate their IMOU, mostly due to disagreements with external stakeholders.

Site received many other shared responsibility "fair" ratings, indicating the difficulties of working together. In particular, communication internal to C^3RS (i.e., between the PRT and Support Team) and external to C^3RS (from the PRT and Support Team to the workforce) received a "fair" rating for all sites. The Support Teams wanted more justification information and analysis from the PRTs concerning their corrective action recommendations. The PRTs

wanted more communication from the Support Teams regarding the status of corrective actions. Labor and management in the field needed to hear more about C^{3}RS accomplishments. Some of the lack of communication to the field was caused by worries over confidentiality; some was due to lack of time to write more publications like newsletters. There were fair ratings in the areas of data analysis (e.g., cost-benefit analysis) and process efficiency (e.g., breaking PRT into subgroups). The training from the Implementation Team was considered sufficient and rated "good" across all four sites. Confidentiality was also rated "good" because sites had trust that the manner in which reports were submitted and processed, and the manner in which discipline was avoided, were consistent with the IMOU and expectations during the demonstrations.

Implementing Corrective Actions were rated "good" for cooperative work among labor, management, PRT, and Support Team for Sites 1, 2, and 3, who worked together to implement actions. However, Site 4 received a rating of "poor" as implementation of corrective actions was mostly left to local managers and labor did not participate. While tracking of corrective actions was rated unfavorably at three of the four sites, accountability was rated "good" at Sites 2 and 3. Site 4 received "poor" ratings on all detailed factors related to this main factor.

Effective Dispute Resolution received a rating of "good" at Sites 1, 2, and 3 because they resolved disputes ensuring confidentiality and trust were maintained. Site 4 received a rating of "poor" due to several disputes regarding when and how discipline protection should be applied for reporters. In those disputes, the reporters were disciplined at the conclusion of the disputes rather than receiving protection from discipline. These resolutions seemed to decrease trust in the program, which resulted in a further reduction in the amount of detail reporters were willing to include, hampering the ability of the PRT to come up with contributing factors and related, effective corrective actions.

Perceived Value received a rating of "good" for Sites 1, 2, and 3. Interviewees at those three sites believed that C^{3}RS had a positive impact on safety culture and safety, and produced effective corrective actions. Interviewees at those three sites also believed that C^{3}RS achieved cost savings as a result of reduced discipline and incidents.

5.3.1.7 Cross Site Sustainability Results

This section discusses the long-term sustainability of C^{3}RS at the four sites after their demonstration periods ended. The next section will discuss the sustainability of C^{3}RS in the railroad industry. Reviewing the ratings on sustainment for the four sites one can see that two of the four sites sustained their efforts post demonstration period. While three of the sites had favorable implementation ratings and observed impacts, one of these sites did not sustain because the internal and external stakeholders could not come to agreement on the revisions needed to the IMOU. This inability to agree on a new IMOU is covered under the implementation factor "Shared Responsibility – Revision to IMOU." The fourth site which also withdrew received a poor rating on this same detailed factor for the same reason— it could not agree with external stakeholders on the revision of the IMOU.

Sites 2 and 3 sustained their involvement in C^{3}RS for several reasons. They deemed their demonstration to be successful and had support from FRA to continue. FRA set up the Human Performance Division to support C^{3}RS going forward and staffed it with multiple personnel. Both labor and management had good support for C^{3}RS, with only minor concerns. Labor and management support were sufficient enough to agree to a long-term C^{3}RS program using NASA as the Third-Party.

Site 1 deemed their demonstration pilot as successful, based on the positive impact on safety and safety culture. They attempted to remain long-term in the C^{3}RS program after their demonstration, but both management and labor had concerns about confidentiality and FRA's long-term plans to share aggregated de-identified close call data with the industry. To qualify for discipline immunity waivers, FRA wanted railroads to participate in the national model so events could be captured in its deidentified database, contributing to data for the safety alert feature of the program. The railroad was concerned that the public would be able to identify them in the database. As of the time of this report, Site 1 was no longer participating in FRA's C^{3}RS. (However, because Site 1 appreciated the value of close call reporting, and interviewees said that, Site 1 later developed an internal safety program to proactively look for causes of incidents without waivers.)

Site 4 did not deem their demonstration pilot successful. They attempted to negotiate their IMOU to overcome some issues with their implementation in the last year of the demonstration pilot, but they did not reach a consensus. For example, there were disagreements about requiring each crew member to report to C^{3}RS receive discipline immunity versus allowing protection to the whole crew based on only one member reporting. As discussed in the implementation section above, the PRT did not feel that that the data being provided by reporters was sufficient to determine the causes of close calls, but national labor was concerned about making whole crew reporting required. Because a consensus to change was not reached and labor and management at the site agreed that continuing in the same manner as the demonstration pilot was not worthwhile, the site withdrew at the end of their demonstration.

5.3.2 Industry Sustainability Results

In this section, analysis of all the interview, project records, and field notes data is presented to assess sustainability at the industry level and used to create a different fishbone diagram with different factors. See Figure 17 for the industry sustainment main factors and definitions. Figure 31 summarizes industry sustainability results. The text below the figure presents details on the ratings for each sustainability factor.

Effects: C^{3}RS Sustained in the Industry. As of this writing, eight railroads were participating in C^{3}RS, with the new additions of Southeastern Pennsylvania Transportation Authority, Strasburg Railroad, Massachusetts Bay Transportation Authority, Metra, Metro-North, and the Long Island Railroad. The two passenger demonstration site railroads were still involved. So in essence, the sustainment of C^{3}RS was almost entirely centered on passenger carriers. Class 1 freight was entirely absent from FRA's C^{3}RS program. Shortlines (Class 2 and 3) freight carriers were almost entirely absent. In light of this situation, understanding industry sustainability needed to be treated separately for passenger versus freight. Because the nature of Class 1 business is so different from that of smaller railroads, it seemed reasonable not to lump Class 1, 2, and 3 railroads together. Figure 31 reflects these different groups, and provides a separate sustainability effect rating for each type of railroad. The Evaluation Team rated the sustainment effect for passenger carriers as "good," for Class 1 railroads "poor," and for shortlines "poor." The factors influencing these effects are described below.

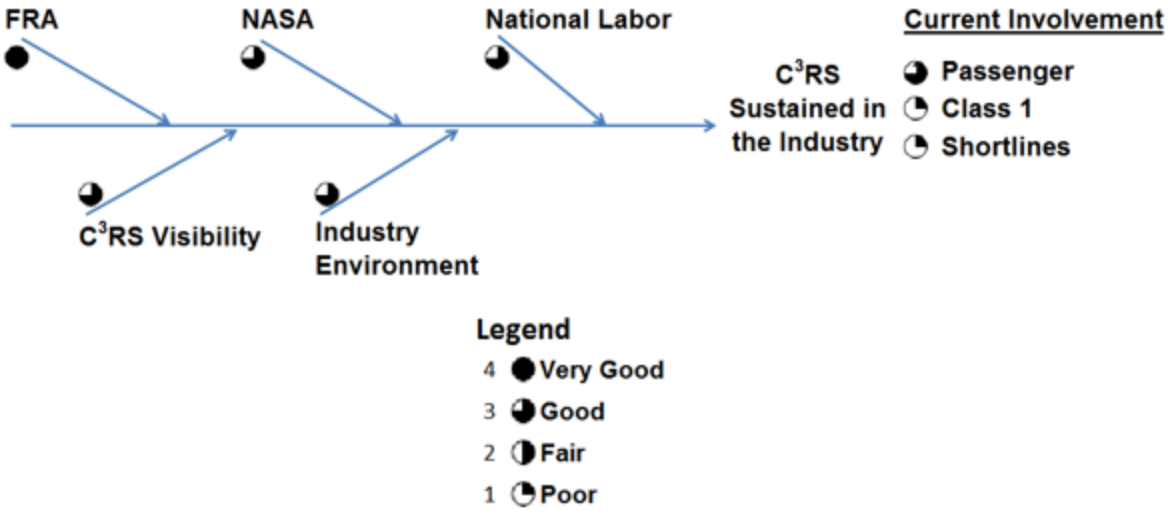


Figure 31. Summary of Industry Sustainability Ratings

FRA – FRA received a rating of "very good" in their support for C^{3}RS sustainability within the U.S. railroad industry. FRA made significant investment and progress to sustain C^{3}RS as an ongoing national program. They shifted C^{3}RS from a research program to a national non-regulatory safety program. The program continued to grow following the demonstration pilots with support from FRA, FRA's funding of NASA, and eight currently participating railroads, two of which were the original pilot sites.

Another significant action FRA took was to establish the Human Performance Division to support C^{3}RS going forward and to staff it with multiple personnel within the RRS, thereby making it part of their ongoing railroad oversight role. These personnel helped new sites negotiate their IMOUs and provided training. FRA selected NASA as the Third-Party and provided funding.

FRA funded the development of a new software tool to support the analysis of event reports and the tracking of actions to fix identified problems. The acquisition involved Peer Review Teams (PRT) when developing requirements.

FRA proposed a regulation for freight railroads to set up Risk Reduction Programs (RRP) with C^{3}RS as a voluntary element. FRA implemented a new regulation for passenger railroads to set up System Safety Programs (SSP), with C^{3}RS as a voluntary element.

FRA also sponsored and organized a yearly C^{3}RS User Group to facilitate the sharing of practices and corrective actions among participating railroads.

Toward the very end of the evaluation, FRA experienced some budget constraints causing them to reduce their funding to the Third-Party. They began to explore ways to reduce costs with minimal detriment to the program. One method that they tried was to reduce the number of Third-Party call backs to reporters. At the 2017 User Group, the participating sites expressed some concerns about the potential impact on data completeness from this change.

NASA – Received a rating of "good" for their support of C^{3}RS sustainability. They continued to provide Third-Party report collection to C^{3}RS as it grew. NASA began to release some information about overall trends in reporting and issued safety alerts. NASA worked on a

process to make the data collection more cost effective and timely in the future by moving to a sampling method for reporter interviews. NASA protected data from disclosure and protected the identities of railroads and individuals. One challenge was that NASA data protection procedures required stripping of identifying data; some PRTs occasionally wished they had more detail to help them determine the contributing causes and create effective corrective actions.

National Labor – Received a rating of "good" for their support of C^{3}RS sustainability. They continued to support labor at sites that implemented C^{3}RS. In addition, engineering and mechanical crafts began to join at multiple sites. While commitment to C^{3}RS is ultimately a management decision, such commitments could only come with strong labor union support. In post demonstration period interviews one national union official mentioned concern about the implication of new technology on C^{3}RS. For instance as inward facing cameras are added to locomotive cabs, carriers may have less motivation for C^{3}RS with this real time surveillance. New and existing C^{3}RS sites will need to consider revising IMOUs to address such arrangements. As seen with the challenges for IMOU revisions with the four demonstration pilots, trust and cooperation will be essential to successfully navigate such changes.

For the list of all unions involved as of the writing of this report see Table 26.

C^{3}RS Visibility – Received a rating of "good" for its support of C^{3}RS sustainability. In 2012 the Evaluation Team, with the approval of the demonstration sites, began to share positive impacts on safety with the industry. These results were used by FRA to justify funding from Congress and to help promote C^{3}RS across the industry. As more railroads joined C^{3}RS, there was a positive feedback loop, increasing FRA management's interest in supporting and funding C^{3}RS and industry's interest (from passenger railroads). Based on the positive experience of Transportation crafts, Mechanical and Engineering crafts joined as well (Table 26).

Industry Environment – Received a rating of "good" for its support of C^{3}RS sustainability. They supported C^{3}RS in part because of several traumatic fatal accidents which occurred in passenger railroads, putting pressure on carriers to take visible steps to improve safety. This had some influence in recruiting passenger railroads in the Northeast Corridor.

Table 26. List of Labor Unions Involved in the C^{3}RS Program

Organization	Craft
Brotherhood of Locomotive Engineers and Trainmen (BLET)	Transportation
Sheet Metal and Rail Transportation – Transportation Division (SMART TD)	Transportation
Association of Commuter Rail Employees (ACRE)	Transportation and Engineering
American Train Dispatchers Association (ATDA)	Transportation
Transportation Communications Union (TCU)	Transportation and Mechanical
International Brotherhood of Electrical Workers (IBEW)	Mechanical and Engineering
American Railway and Airway Supervisors Association (ARASA)	Mechanical and Engineering
National Conference of Fireman and Oilers (NCFO)	Mechanical
Sheet Metal and Rail Transportation – Sheet Metal Division (SMART-SM)	Mechanical
Brotherhood of Railway Signalmen (BRS)	Engineering
Brotherhood of Maintenance of Way Employees (BMWE)	Engineering
Metropolitan Alliance of Police	Engineering
International Association of Machinist and Aerospace Workers (IAMAW)	Mechanical
Transport Workers Union of America (TWU)	Mechanical
International Brotherhood of Teamsters	Engineering

5.4 Railroad Safety Culture Survey

5.4.1 Railroad Safety Culture Survey Scale Results

A 2x3 MANOVA multivariate analysis revealed a significant interaction between date and site (Wilks' Lambda=0.95, $p<0.001$) for the Railroad Safety Culture Scales. Recall that the "date" variable had values of baseline or final. The "site" variable had values of Site 1, Site 2, or Site 4. The 2x3 ANOVA revealed that most scales contained significant interaction between date and site, except for Organizational Concern for Employees and Work-Safety Priorities (Table 27).

Table 27. 2x3 ANOVA Results for Railroad Safety Culture Survey Scales

	Date	**Site**	**Date*Site**
Scale Name	***p***	***p***	***p***
Organizational Concern for Employees	0.15	0.38	0.93
Labor-Management Relations	0.00	0.00	0.00
Management Safety	0.15	0.00	0.00
Supervisor Fairness	0.00	0.00	0.00
Supervisor-Employee Relationships	0.00	0.00	0.00
Raising Concerns with Supervisors	0.00	0.02	0.04
Work-Safety Priorities	0.00	0.00	0.83
Helping Behavior	0.01	0.00	0.00
Coworker Safety	0.00	0.00	0.00

Next, for each site, the multivariate analyses, with date as the independent variable and the Railroad Safety Culture Scales averages as dependent variables, showed that all three sites had significant changes over time (Site 1 Wilks' Lambda=0.84, $p<0.001$, Site 2 Wilks' Lambda =0.96, $p<0.001$, and Site 4 Wilks' Lambda=0.74, $p<0.001$). The following one-way ANOVA tests revealed, for each site, which scales changed significantly. The scale means and p-values are shown in Table 28 and discussed below for each scale. In the table, light gray shading shows significant increases and dark gray shading shows significant decreases.

Finally, the multivariate test for each pair of sites, with date and site as independent variables and the Railroad Safety Culture scale averages as dependent variables, revealed that significant differences in amount of change over time between the sites (Site 1 x Site 2 Wilks' Lambda=0.97, $p<0.001$, Site 1 x Site 4 Wilks' Lambda=0.95, $p<0.001$, Site 2 x Site 4 Wilks' Lambda=0.96, $p<0.001$). The 2x2 ANOVA revealed which scales had different changes between pairs of sites. The differences are discussed below for each scale.

Table 28. One-Way ANOVA Results per Site for Railroad Safety Culture Survey Questions

	Site 1			Site 2			Site 4		
Scale Name	**Baseline Mean**	**Final Mean**	***p***	**Baseline Mean**	**Final Mean**	***p***	**Baseline Mean**	**Final Mean**	***p***
Organizational Concern for Employees	2.2	2.2	0.17	2.2	2.2	0.12	2.2	2.4	0.19
Labor-Management Relations	2.2	2.6	0.00	2.6	2.7	0.01	2.5	2.6	0.25
Management Safety	2.8	3.0	0.00	2.8	2.8	0.55	3.2	3.2	0.77
Supervisor Fairness	2.4	2.8	0.00	2.6	2.8	0.00	2.7	3.2	0.00
Supervisor-Employee Relationships	2.7	3.1	0.00	2.9	3.1	0.00	3.1	3.6	0.00
Raising Concerns with Supervisors	3.4	3.6	0.00	3.5	3.5	0.09	3.6	3.8	0.01
Work-Safety Priorities	3.7	3.8	0.00	3.5	3.6	0.00	3.7	3.9	0.04
Helping Behavior	3.5	3.5	0.38	3.4	3.5	0.00	3.4	2.9	0.00
Coworker Safety	4.1	3.9	0.00	3.8	3.8	0.08	3.9	3.8	0.08

5.4.1.1 Organizational/ Managerial Scales

Three Organizational/ Managerial scales were available for analysis: Organization Concern for Employees (OCE), Labor/Management Relations (LMR), and Management Safety (SafeM). The means for all sites for these three scales appear in Figure 32. (Site 3 did not have a related scale for LMR.)

For Organizational Concern for Employees, the 2x3 ANOVA revealed no significant interaction between date and site (p=0.93) and no significant main effect of date (p=0.15). OCE did not significantly change for the Sites 1, 2, or 4. In contrast, Site 3 did increase in their OCE-related scale. (Recall that Site 3 performed their own survey, and the C^3RS Evaluation Team did not receive data on statistical significance.)

For Labor-Management Relations, the 2x3 ANOVA revealed significant interaction between date and site (p<0.001). The one-way ANOVA with date as independent variable for each site showed Sites 1 and 2 significantly improved and Site 4 did not change significantly (Site 1 p<0.001, Site 2 p=0.01, Site 4 p=0.25). The 2x2 ANOVA tests showed that Site 1 changed significantly more than Sites 2 and 4 (p<0.001 and p=0.01, respectively). The 2x2 also showed that Site 2 did not change more than Site 4 (p=0.79). Site 3 did not have a related scale.

For Management Safety, the 2x3 ANOVA revealed significant interaction between date and site (p<0.001). The one-way ANOVA for each site showed Site 1 significantly improved and Site 2 and Site 4 did not change significantly (Site 1 p<0.001, Site 2 p=0.55, Site 4 p=0.77). The 2x2 ANOVA tests showed that Site 1 changed significantly more than Site 2 and Site 4 (p<0.001 and p=0.05, respectively). The 2x2 also showed that Site 2 and 4 did not change significantly differently from each other (p=0.86). In comparison, Site 3 increased in their Management Safety-related scale.

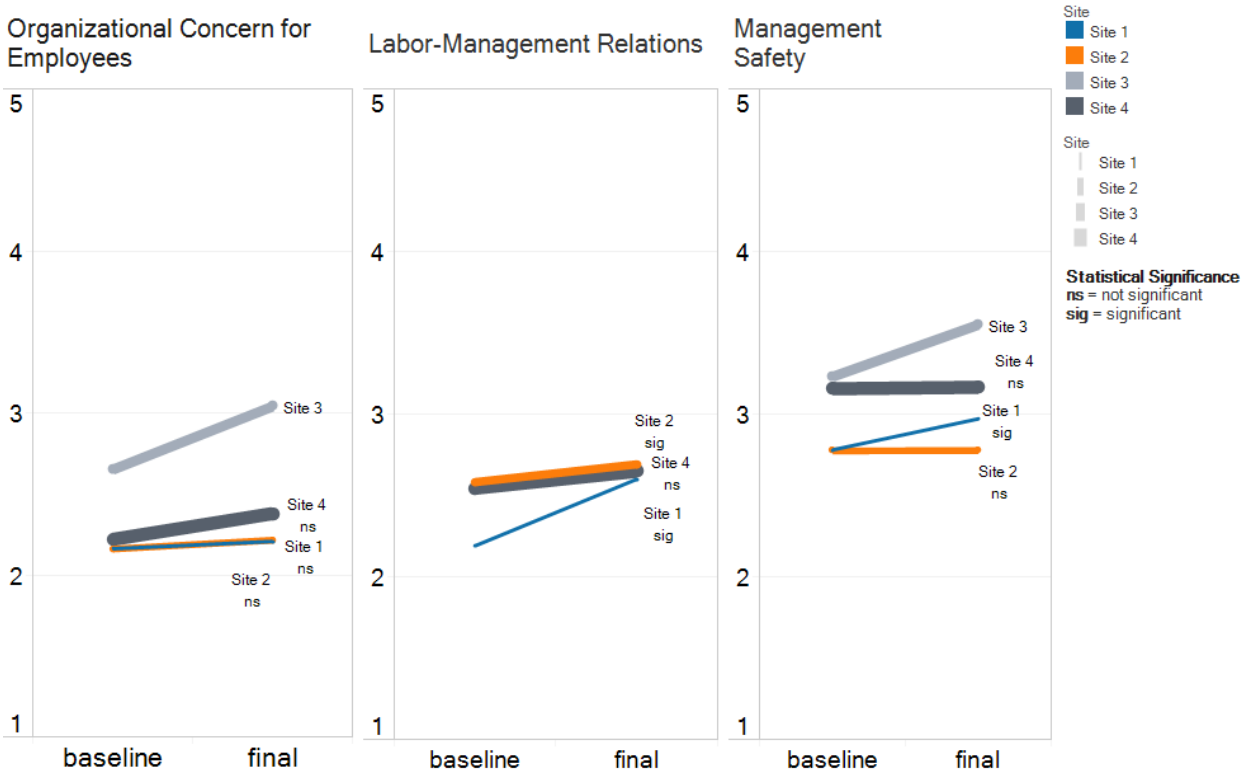


Figure 32. Organizational/Management Safety Culture Scale Values – Comparing Labor's Views at Baseline vs. Final[34]

5.4.1.2 Supervisor Scales

Three Supervisor scales were available for analysis: Supervisor Fairness (SF), Supervisor-Employee Relationships (SER), and Raising Concerns with Supervisors (RCS). The means for all sites for these three scales appear in Figure 33.

For Supervisor Fairness, the 2x3 ANOVA revealed significant interaction between date and site (p<0.001). The one-way ANOVA for each site showed that all three sites significantly improved (Site 1 p<0.001, Site 2 p<0.001, Site 4 p<0.001). The 2x2 ANOVA tests showed that Site 1 and Site 4 did not change significantly differently from each other (p=0.73). The 2x2 also showed that Site 2 improved significantly less than both Site 1 and Site 4 (p<0.001 and p=0.01, respectively). In comparison, Site 3 also increased in their Supervisor Fairness-related scale.

For Supervisor-Employee Relationships, the 2x3 ANOVA revealed significant interaction between date and site (p<0.001). The one-way ANOVA for each site showed that all three sites significantly improved (Site 1 p<0.001, Site 2 p<0.001, Site 4 p<0.001). The 2x2 ANOVA tests showed that Site 1 and Site 4 did not change significantly differently from each other (p=0.44). The 2x2 also showed that Site 2 improved significantly less than Site 1 and 4 (p<0.001 and p=0.02, respectively). In comparison, Site 3 also increased in their Supervisor-Employee Relationships-related scale.

[34] Legend: Statistically significant change from baseline to final = "sig," Not significant="ns"

For Raising Concerns with Supervisors, the 2x3 ANOVA revealed significant interaction between date and site (p=0.04). The one-way ANOVA for each site showed that Sites 1 and 4 significantly improved, and Site 2 did not change significantly (Site 1 p<0.001, Site 2 p=0.09, Site 4 p=0.01). The 2x2 ANOVA tests showed that Site 1 changed significantly more than Site 2 (p=0.02). The 2x2 also showed that Site 1 and 4 and Sites 2 and 4 did not change significantly differently from each other (p=0.75and p=0.13, respectively). In comparison, Site 3 increased in their Raising Concerns with Supervisors-related scale.

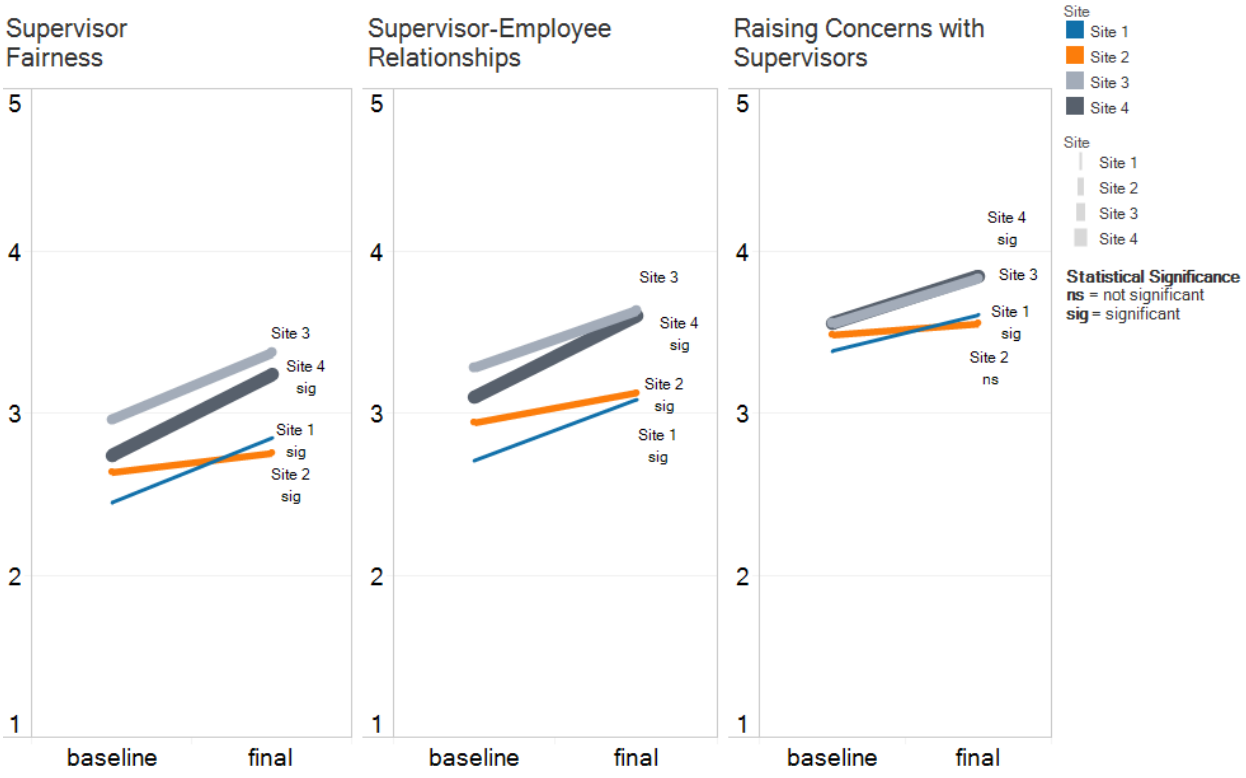


Figure 33. Supervisor Safety Culture Scale Values – Comparing Labor's Views at Baseline vs. Final

5.4.1.3 Coworker Scales

Three Coworker scales were available for analysis: Work Safety Priorities (WSP), Helping Behavior (Help), and Coworker. All the scales had high values at baseline, leaving little room for improvement. The means for all sites for these three scales are in Figure 34. (Site 3 did not have a related scale for WSP or Helping Behavior.)

For Work-Safety Priorities, the 2x3 ANOVA revealed no significant interaction between date and site (p=0.83), but showed a significant impact from date (p<0.001). The one-way ANOVA for each site showed that all three sites significantly improved (Site 1 p<0.001, Site 2 p<0.001, Site 4 p=0.04). The 2x2 ANOVA tests showed that none of the pairs of sites changed significantly differently from each other (p=0.64, p=0.82, p= 0.65). Site 3 did not have a related scale.

For Helping Behavior, the 2x3 ANOVA revealed significant interaction between date and site (p=0.01). The one-way ANOVA for each site showed that Site 1 did not change significantly, and Site 2 improved significantly, and Site 4 declined significantly (Site 1 p=0.38, Site 2 p<0.001, Site 4 p<0.001). The 2x2 ANOVA tests showed that Site 2 did not change significantly more than Site 1 (p=0.19). The 2x2 also showed that Site 4 declined significantly more than Site 1 and Site 2 (p<0.001 and p<0.001, respectively). Site 3 did not have a related scale.

For Coworker Safety, the 2x3 ANOVA revealed significant interaction between date and site (p<0.001). The one-way ANOVA for each site showed that Site 1 decreased significantly, and Site 2 and Site 4 did not change significantly (Site 1 p<0.001, Site 2 p=0.08, Site 4 p=0.08). The 2x2 ANOVA showed that Site 1 declined significantly more than Site 2 but not significantly differently from Site 4 (p<0.001 and p=0.85, respectively). The 2x2 ANOVA also showed that Site 2 and Site 4 changed significantly differently from each other (p=0.05). In contrast, Site 3 improved in their Coworker-related scale.

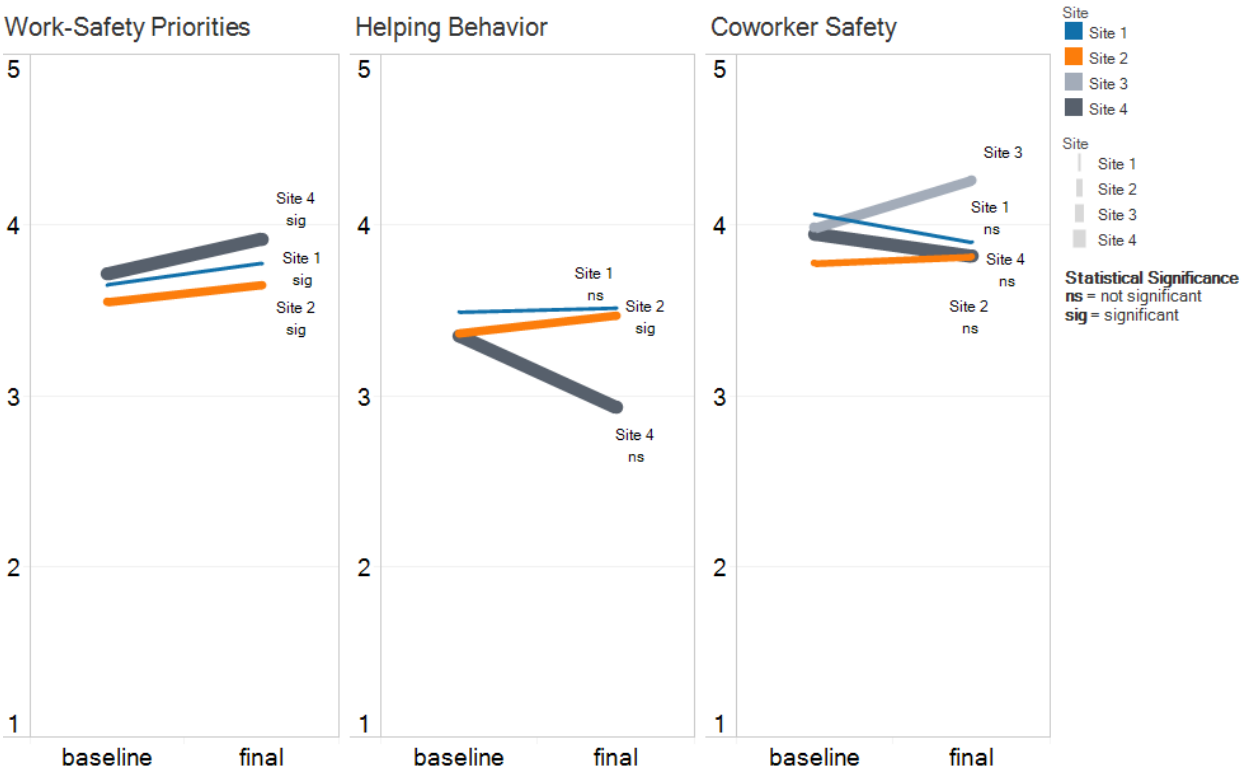


Figure 34. Coworker Safety Culture Scale Values – Comparing Labor's Views at Baseline vs. Final

5.4.2 Railroad Safety Culture Survey Scale – Safe Behavior Results

Four safe behavior scales were available for analysis: Road Crews – Cab Red Zone, Yard Crews – Switching, Dispatching, and Safety Briefings. Only Sites 2 and 4 were surveyed about safe behaviors at baseline and final, allowing for a comparison over time. Only Site 2 was asked about Dispatching behaviors.

The 2x2 ANOVA for each scale revealed that Safety Briefings contained significant interaction between date and site (p<0.001), but Road and Yard behaviors did not (p=0.12 and p=0.15,

respectively). However, the main effect of date, across the two sites, was significant for both Road and Yard behaviors (p=0.01 for each). All these p-values were significant at the Bonferroni-corrected level of 0.0167. Averaged across Sites 2 and 4, Yard behaviors improved significantly from 4.5 at baseline to 4.7 at final, and Road behaviors improved significantly from 3.1 to 3.3.

For Dispatching behaviors they surveyed only Site 2; its t-test showed significant improvement (baseline=4.3, final=4.5, p=0.03).

For Safety Briefings, one-way ANOVA for each site showed that Site 2 improved significantly (baseline=3.3, final=3.6, p<0.001), and Site 4 did not change significantly (baseline=3.9, final=3.8, p=0.18). The 2x2 ANOVA showed that Site 2 improved significantly more than Site 4 (p<0.001).

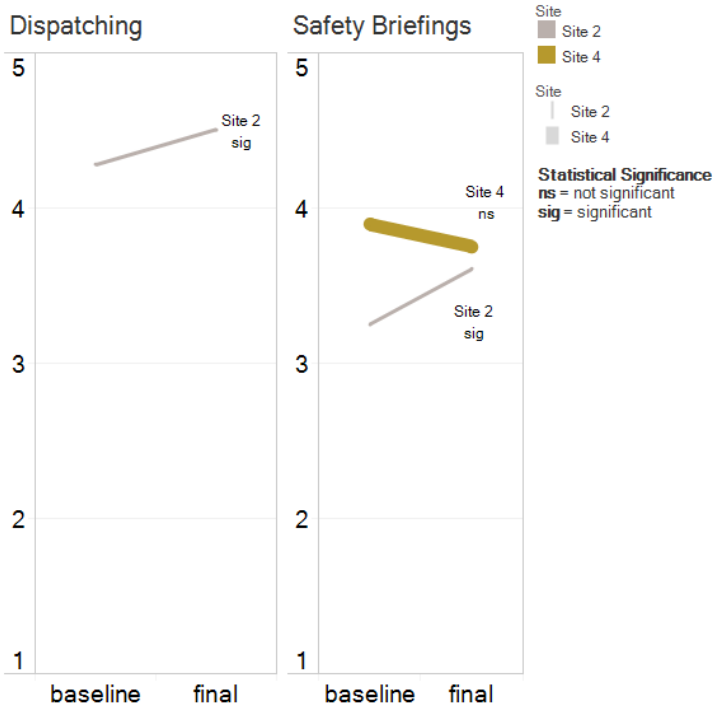


Figure 35. Safe Behaviors Scale Values – Comparing Labor's Views at Baseline vs. Final

5.4.3 Impact and Sustainability Survey Questions

Three questions about C^{3}RS and its impact were available for analysis. Only Sites 1 and 4 were surveyed about impact at baseline and final, allowing for a comparison over time. One question about sustainability was available for analysis; it was answered by Sites 1 and 2.

2x2 MANOVA multivariate analysis revealed a significant interaction between date and site (Wilks' Lambda=0.99, p<0.01) for impact questions about C^{3}RS increasing awareness of safety issues within labor, increasing awareness of safety issues within management, and resulting in improved safety. The 2x2 ANOVA showed all three questions had significant interaction between date and site (p=0.01, p<0.001, p=0.01, respectively).

For increased awareness of safety issues within labor, one-way ANOVAs for each site showed that Site 1 and Site 4 improved significantly (p<0.001 and p=0.02). The 2x2 ANOVA showed that Site 1 improved more than Site 4 (p=0.01).

For increased awareness of safety issues with management, one-way ANOVAs for each site showed that Site 1 improved significantly (p<0.001), and Site 4 did not change significantly (p=0.20). The 2x2 ANOVA showed that Site 1 improved more than Site 4 (p<0.01).

For C^{3}RS resulting in improved safety, one-way ANOVAs for each site showed that Site 1 and Site 4 improved significantly (p<0.001 and p<0.001). The 2x2 ANOVA showed that Site 1 improved more than Site 4 (p=0.01).

For all three impact questions, Site 2 only had final values and showed that on average respondents agreed that C^{3}RS had a positive impact (95% confidence intervals were above "3." i.e., above "neutral" (Improved Safety (3.30 to 3.40), Labor Awareness (3.50 to 3.60), and Management Awareness (3.26 to 3.37)). Site 1 and Site 4's average respondents at final also agreed that C^{3}RS had a positive impact (95% confidence intervals for Improved Safety at Site 1 (3.6 to 3.7) and Site 4 (3.3 to 3.7)).

The final question in this section asked if the respondent thought C^{3}RS would remain in operation for the foreseeable future. One-way ANOVAs for each site showed that both Site 1 and Site 2 improved significantly (p<0.001 for both). The 2x2 ANOVA showed that Site 1 had a bigger improvement than Site 2 (p<0.001).

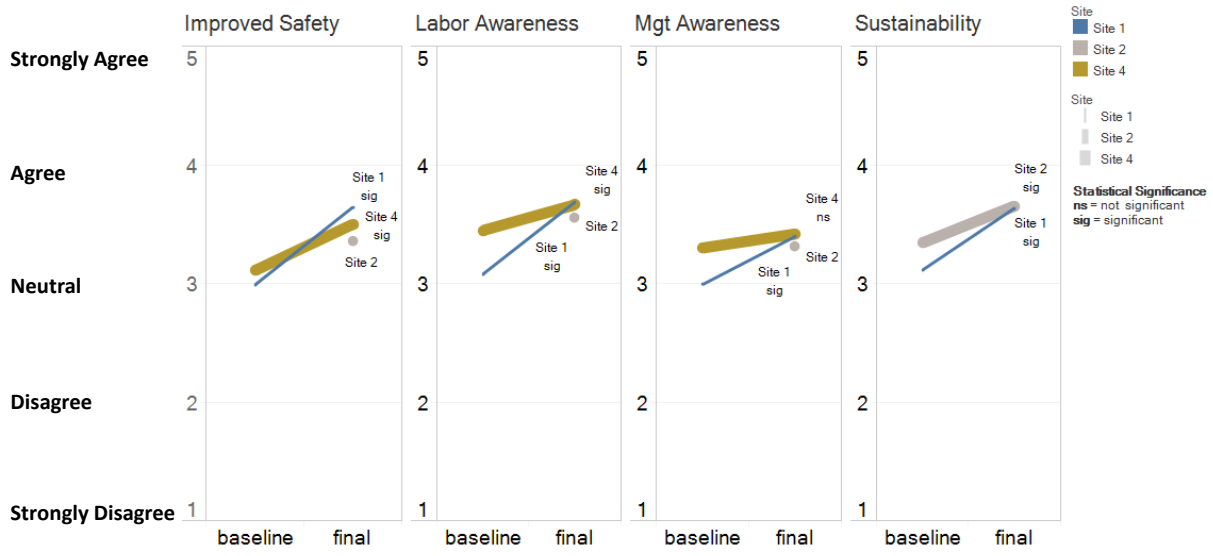


Figure 36. C^{3}RS Impact and Sustainability Scale Values – Comparing Labor's Views at Baseline vs. Final

5.4.4 C^{3}RS Reporting Questions

When the survey asked respondents if they were willing to submit a report to C^{3}RS, Sites 1 and 2 significantly increased from half of respondents being willing to submit a report to C^{3}RS at baseline to 83% at final (Site 1 Chi-squared=83.12, p<0.001 and Site 2 Chi-squared=86.26, p<0.001).

Site 4 significantly decreased from 82% to 62% of respondents being willing to submit a report (Chi-squared=8.08, $p<0.001$). When asked "why not," respondents in Site 4 said that the process was too much of a bother, they were worried about being punished, and they did not think it would result in any changes (percent of respondents at final who selected that reason were 55%, 29%, and 23%, respectively).

5.4.5 Summary of Survey Results

The C^{3}RS Railroad Safety Culture Survey showed that labor respondents had significantly improved perceptions of safety culture, behaviors, and the C^{3}RS program (Table 29). In this summary, the results from Site 3's survey were included as "significant." While the data were not available to verify whether or not Site 3 improved statistically significantly, seeing the similarities between their increases and those of the other three sites was informative.

All four sites had significant improvements in the Supervisor safety scales. Site 1 and Site 3 significantly improved in some of the Organizational/Managerial scales. Site 2 and 3 significantly improved in the coworkers scales. Site 2 significantly improved in the Safe Behavior scales. Sites 1, 2, and 4 believed the C^{3}RS improved safety. Sites 1 and 2's respondents indicated a significantly greater willingness to report to C^{3}RS.

Table 29. Summary of Changes in Labor's Perceptions of Safety Culture

	Site 1	Site 2	Site 3	Site 4
Organizational/Managerial	↑	-	↑	-
Supervisor	↑	↑	↑	↑
Coworker	-	↑	↑	-
Safe Behaviors	?	↑	?	-
Impact from C^{3}RS	↑	↑	?	↑
Willingness to Report to C^{3}RS	↑	↑	?	-

Legend:
↑ = mostly improvement
\- = mixed results/ little to no change
? = no data

In Table 29 up arrows indicate "mostly improvement" and were developed using this criteria:

- **Organizational/Managerial:** An up arrow indicates that at least two of the three scales improved significantly (except for Site 3, whose significance level is unknown, but whose scale increased).
- **Supervisor:** An up arrow indicates that at least two of the three scales improved significantly (except for Site 3, whose significance level is unknown, but whose scale increased).
- **Coworker**: An up arrow indicates that at least two of the three scales improved significantly (except for Site 3, which had only one Coworker–related scale and whose significance level is unknown, but the scale increased).

- **Safe Behaviors**: An up arrow indicates that all of the applicable scales significantly increased.
- **Impact from C^{3}RS**: An up arrow indicates that at final, respondents on average agreed that C^{3}RS resulted in improved safety.
- **Willingness to Report to C^{3}RS**: An up arrow indicates a significant increase in the % of respondents who said they were willing to submit a report to C^{3}RS.

Figure 37 shows the summary of the Organizational/Managerial, Supervisor, and Coworker Railroad Safety Culture Scales. In the figure, baseline values for a scale are indicated by an open circle. Final values are shown with a shaded symbol: an "up" arrow if the final was statistically significantly higher than baseline; a solid circle if the final value was not statistically different; and a "down" arrow if the final was statistically significantly lower than baseline. (Site 3 has a differently shaped "up" arrow because the Evaluation Team did not receive the data to be able to perform statistical tests to determine if the increases were significant.)

The most improvements in the Railroad Safety Culture Scales were in the scales related to employees' relationships with supervisors. At all four sites, labor perceived that supervisors exhibited significantly more fairness and had improved relationships with them. At three of the sites, employees felt more comfortable raising concerns with supervisors.

Sites 1, 2, and 3 had some improvements in the Organizational/Managerial scales, related to their perceptions of senior and middle management's relationship and concern for labor (significant at Site 1 and 2). Site 1 felt that labor-management relationships and management's commitment to safety had improved. Site 2 only had a small increase in the perception of labor-management relationships. (Unrelated to C^{3}RS, Site 2 also had some tension in labor-management relations over contracts occurring during the final survey.) Site 3 had improvements in perceptions of the organization's concern for labor and management's commitment to safety. Site 4 had no changes in their Organizational/Managerial scales.

Sites 1 and 4 each observed a significant decrease in one Coworker safety culture scale. Site 1 showed a small decrease in their view of coworker's commitment to safety, but the overall views of coworkers were still high. Site 4 had a larger reduced perception of the amount of helping behavior among labor.

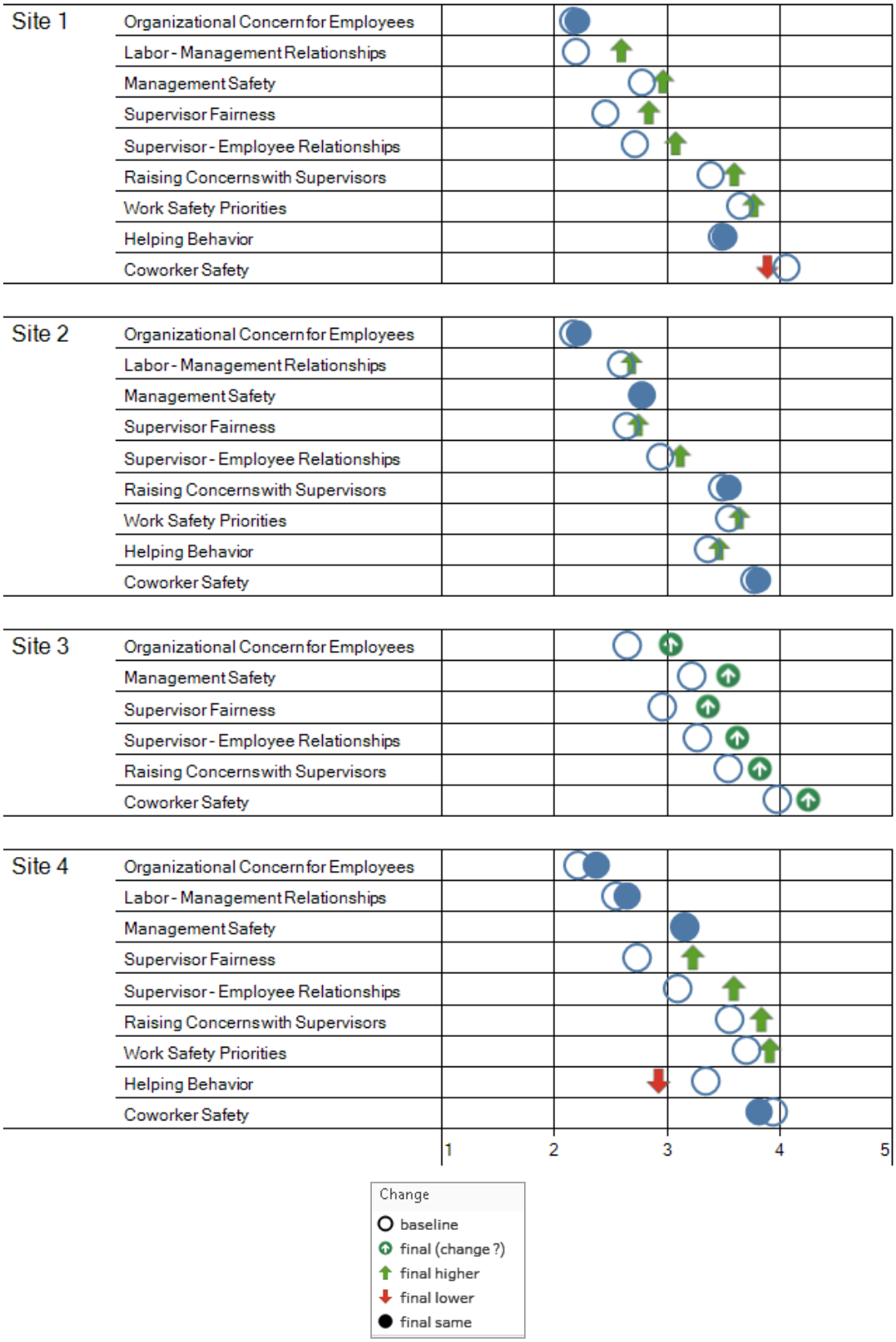


Figure 37. Safety Scale Results across Sites

5.5 Corporate Archival Data

The Evaluation Team saw that corrective actions had indeed been implemented, justifying a quantitative analysis on the impact of C^{3}RS (Patton, 1997). See below for results from each site; a summary appears at the end in Section 5.5.5.

5.5.1 Corporate Data Results at Site 1

5.5.1.1 Derailments at Site 1

The Evaluation Team examined human factor derailments per 100,000 cars moved at Site 1 and at a comparison site at the same railroad, including both FRA reportable and non-reportable derailments. The team selected this metric for several reasons. Site 1 collected many close call reports related to derailments and implemented related corrective actions, e.g., adding a camera to the yard, initiating a new policy assuring derail removal, and introducing reflective switch targets to make switch positions more visible. Interviewees said that derailments were decreasing from those actions. Site 1's management believed that C^{3}RS influenced only human-factors derailments, so they did not provide the Evaluation Team with non-human factors incident data.

Site 1's rate of human factors derailments significantly decreased 41% from before to after C^{3}RS corrective actions started (from 4.85 to 2.87 derailments per 100,000 car-moves). A Poisson rate comparison of the before and after rates had the *F* statistic $F(276,410) = 1.68$, $p<0.001$ (see Table 30, Figure 38, Figure 39). The comparison site did not show a significant change (from 2.15 to 2.35 derailments per 100,000 car-moves, $F(322,392) = 1.08$, $p=0.22$).

Table 30. Site 1 Human Factors Derailment Results

Phase	Statistical Metric	C^{3}RS Site Observed Value	Comparison Site Observed Value
Before C^{3}RS corrective actions (3 years before C^{3}RS started, and 6 months during C^{3}RS)	Average Number of Human Factors Derailments per 100K Cars moved	**4.85**	2.15
After C^{3}RS corrective actions (4.75 years)	Average Number of Human Factors Derailments per 100K Cars moved	**2.87**	2.35
	Percent Change	**-41%**	+9%
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**<0.000**	0.224

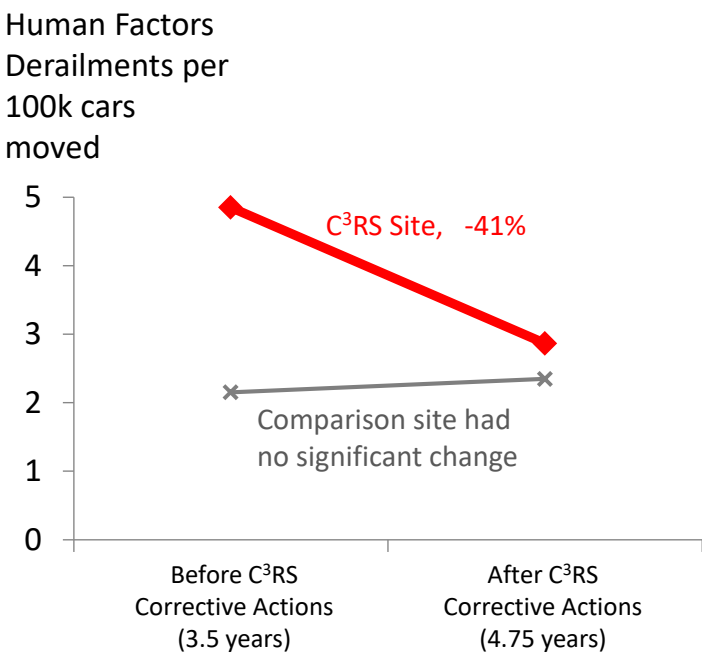


Figure 38. 41% Decrease in Human Factors Derailments at Site 1

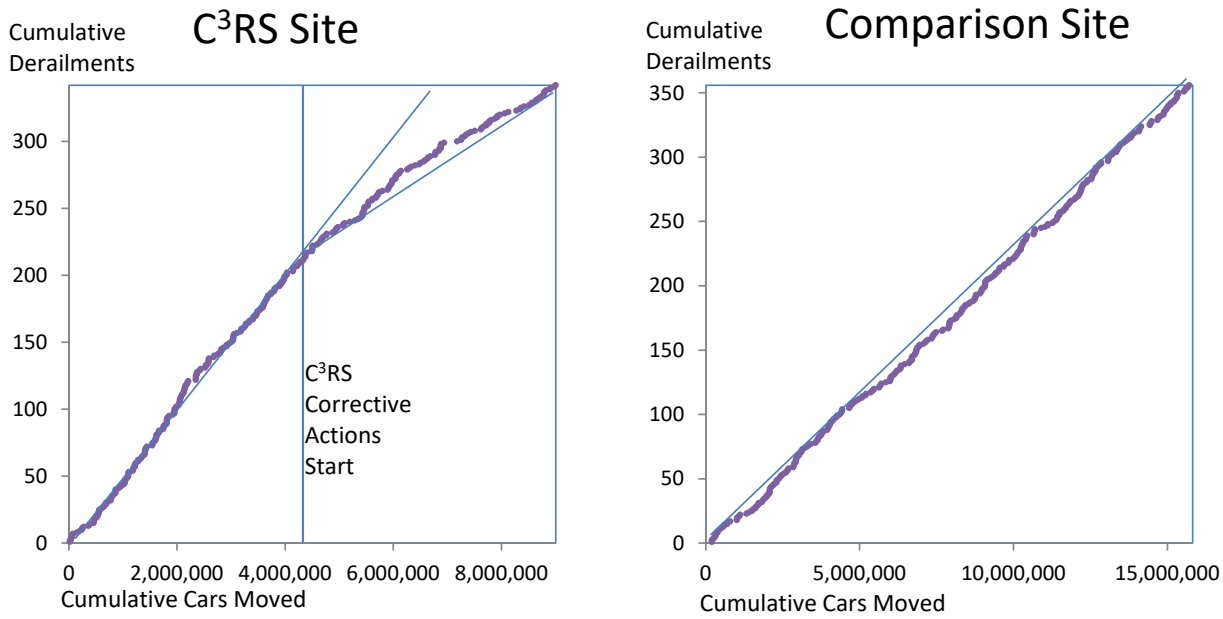


Figure 39. Cumulative Derailments at Site 1[35]

5.5.1.2 Tracking of Run-through Switches and Incident Cost

Knowledge of close calls and precursor incidents appears to help prevent expensive incidents in some cases. For example, knowledge of a run-through switch enables the railroad to fix it

[35] An explanation for how to read cumulative charts is in Section 4.9.5.

quickly and prevent a more costly derailment, even without implementing longer-term corrective actions. Interviewees said that C^{3}RS contributed to better tracking of run-through switches. Looking at the reporting rates, the Evaluation Team found that about 1.75 years into C^{3}RS, run-through switch tracking started to increase, with more than eight times as many run-throughs reported to management and recorded in their safety database. Analysis afterward showed a 53% decrease in the cost of human factors incidents per cars moved (from the time before the run-through switches were being tracked to afterward). To calculate the decrease in cost, all costs were adjusted for inflation.[36] Then the sum of human factors incident costs was calculated for each month. Incidents with costs above the FRA reporting threshold and smaller non-reportable incidents were included. One month it was excluded because it was determined to be an outlier with a single derailment that caused over $2 million in damage. Then a t-test was performed on the natural logarithm (LN) of the monthly costs (two-tailed p= 0.10). While not statistically significant, the Evaluation Team deemed it an intriguing possibility but not a convincing case (Table 31, Figure 40).

Table 31. Site 1 Run-Through Switch Tracking and Human Factors Incident Cost

Phase	Statistical Metric	C^{3}RS Site Observed Value	Statistical Metric	C^{3}RS Site Observed Value
Before and early C^{3}RS (3 years before C^{3}RS started, and 2.75 years during C^{3}RS)	Average number of Human Factors Incidents per month in FRA Category = Other (primarily run-through switches)	0.3	Average Monthly Cost of Human Factors Incidents per 100K cars moved	**$103,131**
Later C^{3}RS (2.3 years)	Average number of Human Factors Incidents per month in FRA Category = Other (primarily run-through switches)	2.8	Average Monthly Cost of Human Factors Incidents per 100K cars moved	**$48,738**
	Percent change	+841%	Percent Change	**-53%**
			p-value (two tailed, t-test on LN of monthly cost)	**0.10**

[36] The Evaluation Team used the Bureau of Labor Statistics Producer Price Index Industry Data for product type railroad rolling stock manufacturing.

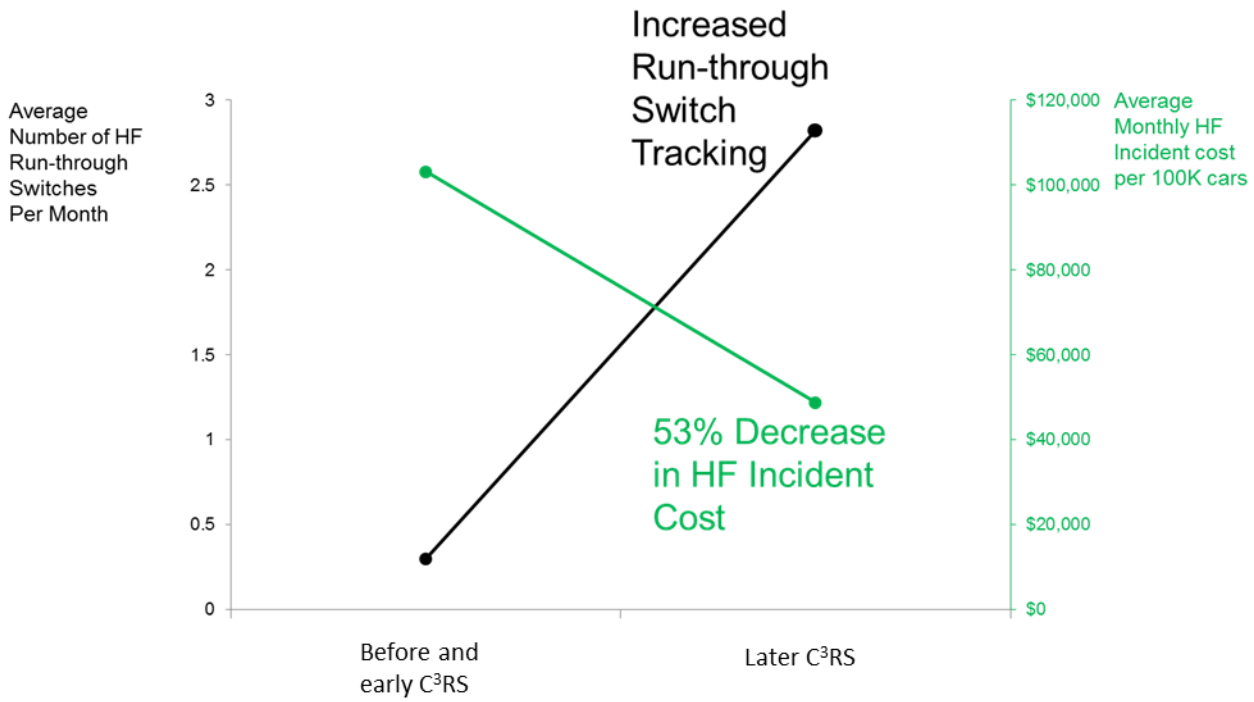


Figure 40. Site 1 Decrease in Human Factors Incident Cost

5.5.1.3 Metrics to Determine Excess Speed

Unfortunately, the Evaluation Team was not able to find a reliable speed-related outcome measure even though there were many reports and corrective actions addressing this issue. In a past study, engineer decertifications were used as a measure to determine if speed-related issues were being constructively addressed (Zuschlag, 2012). However, in C^{3}RS, when a report was submitted the reporter received indemnity and therefore a decertification was not issued. As a result the number of decertifications could be reduced just by the reporting process itself and was therefore found inadequate for an impact measurement.

5.5.2 Corporate Data Results at Site 2

5.5.2.1 Derailments at Site 2

At Site 2, derailments caused by previously run-through switches were the biggest concern. Interviewees agreed that the types of derailments caused by previously run-through switches had decreased at Site 2. The discipline protection from C^{3}RS allowed employees to report run-through switches to their managers without concern. In interviewees said that in the past crews would sometimes try to fix a damaged switch themselves without telling managers and would not always fix it correctly, so the next train through would have a derailment. With C^{3}RS, the reporter told management directly about a damaged switch who would then fix it immediately, thus preventing a future derailment. The PRTs also used the information from the C^{3}RS reports to create corrective actions intended to prevent run-through switches from occurring in the first place; however, the PRT was not sure if those corrective actions were effective because run-through switches continued to occur.

Examination of the reporting data and interviews revealed that Site 2 faced some challenges that decreased reporting to C^{3}RS during "low use" times. For the derailment analysis, "low usage" of C^{3}RS was defined as the time before C^{3}RS started (almost 4 years of data) and 7 months without a run-through-switch report. "High usage" was defined as any month with at least one run-through switch report submitted to C^{3}RS (Figure 41). The timeline below the bar chart in Figure 41 shows when the low usage occurred; it did not occur in consecutive months. This definition was used because the interviewees attributed the reduction in derailments to C^{3}RS enabling employees to tell managers about run-through switches.

Interviewees also told the Evaluation Team that there had been unusually large amounts of snowfall during the years that they were involved in C^{3}RS, which could cause additional derailments. The Evaluation Team decided to use snowfall data from the National Oceanic and Atmospheric Administration (NOAA) as a covariate in a Poisson regression analysis of the derailment rates.

The Evaluation Team analyzed Site 2's derailment data as a Poisson regression analysis with the following model:

$$\text{Ln}(\mu_d) = B_0 + \text{Ln}(s) + B_1 u + B_2 su$$

Where:

μ_d = mean derailments

s = snowfall for that month

u = C^{3}RS usage for that month (1 = high, 0 = low).

That is, the dependent variable was the predicted mean derailments caused by run-through switches for the conditions of each month. The log of each month's snowfall (in millimeters with a small constant included) was fixed with a parameter of 1 to statistically control for the amount of snowfall. The independent variables were an indicator variable of C^{3}RS usage that indicated if there was at least one run-through switch reported to C^{3}RS that month (1=at least one, 0=none) and the interaction of C^{3}RS usage and snowfall.

The regression model was significant (Chi-Square=11.58, p=0.01) and yielded the following parameters:

$$\text{Ln}(\mu_d) = -0.248 + \text{Ln}(s) - 0.702u + 0.004su$$

The analysis found that C^{3}RS use was associated with significantly fewer derailments (B_1= -0.702, p = 0.02). The parameter for the interaction, B_2=0.004, was not significant, p= 0.40. In raw numbers, Site 2's derailments per month during high C^{3}RS usage were 30% percent lower than during low C^{3}RS usage. The regression coefficient of -0.702 indicates that if the snowfall had not been worse during the C^{3}RS years then the derailment reduction would be close to 50%.

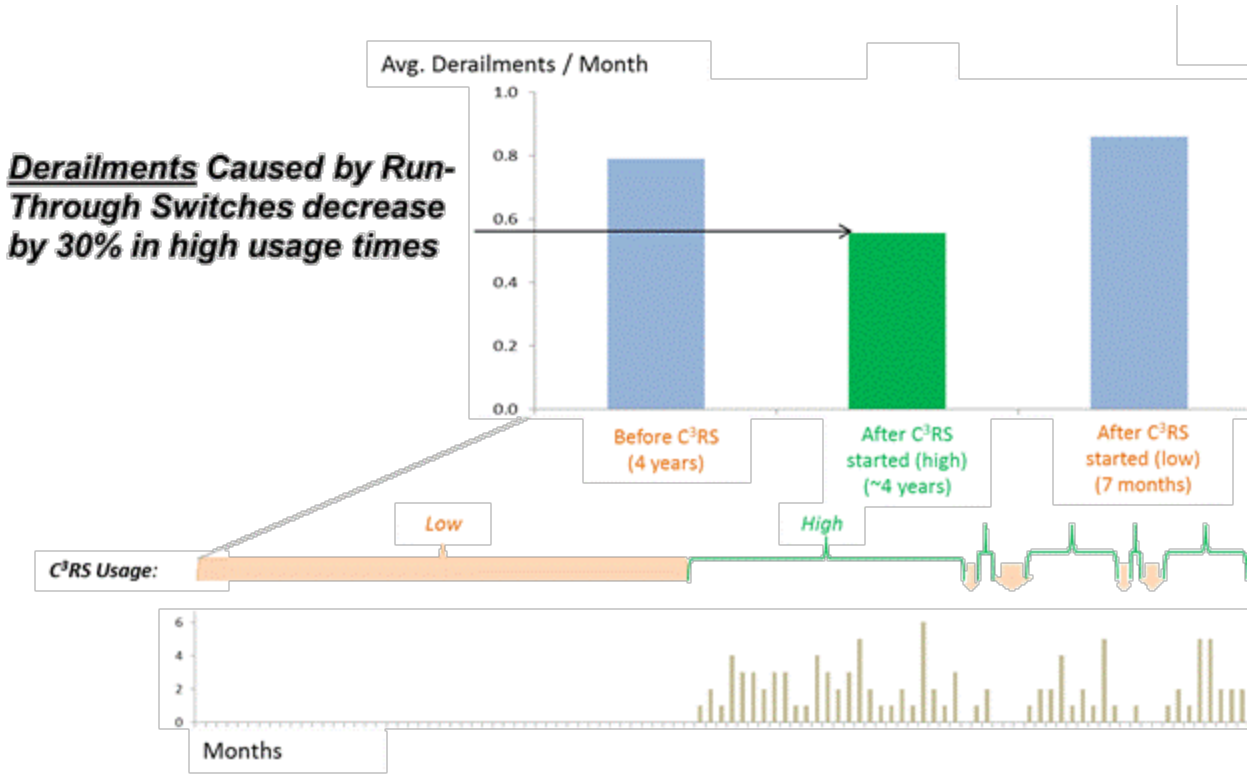


Figure 41. Derailments Decrease 30% at Site 2 during High Usage Times

Table 32. Site 2 Derailment Results

Phase	**Statistical Metric**	**C³RS Site Observed Value**
Low C³RS Usage (3.9 years before C³RS and 7 months of low usage that were not consecutive)	Average Number of Derailments Caused by Run-Through Switches per Month	0.80
High C³RS Usage (3.75 years)	Average Number of Derailments Caused by Run-Through Switches per Month	0.56
	Percent Change	-30%
	p-value (Poisson Regression (Nelson, 2003))	**0.022**

5.5.2.2 Injuries at Site 2

Transportation injuries at Site 2 followed a similar pattern as derailments, with decreases occurring during "high usage" and returning to old values during "low usage." For injuries, they defined "low usage" as time when corrective actions that could reasonably be associated with injuries were not being implemented. For injuries, the low usage time occurred in consecutive months and lasted a little over a year. One major reason that corrective action activity slowed was that the PRT had a large turnover of members all at once who took a while to learn the PRT processes and start being productive. Representatives from Site 2's PRT worked with the Evaluation Team to determine which corrective actions they believed could have an impact on injuries. Then analysis showed that Site 2's injuries per month significantly decreased by 18% when comparing the time low C^{3}RS usage to the high C^{3}RS usage (Figure 42, Table 10) ($F(530,1170) = 1.22$, $p<0.01$, using an F-test for the ratio of two Poisson occurrence rates, Nelson, 2003).

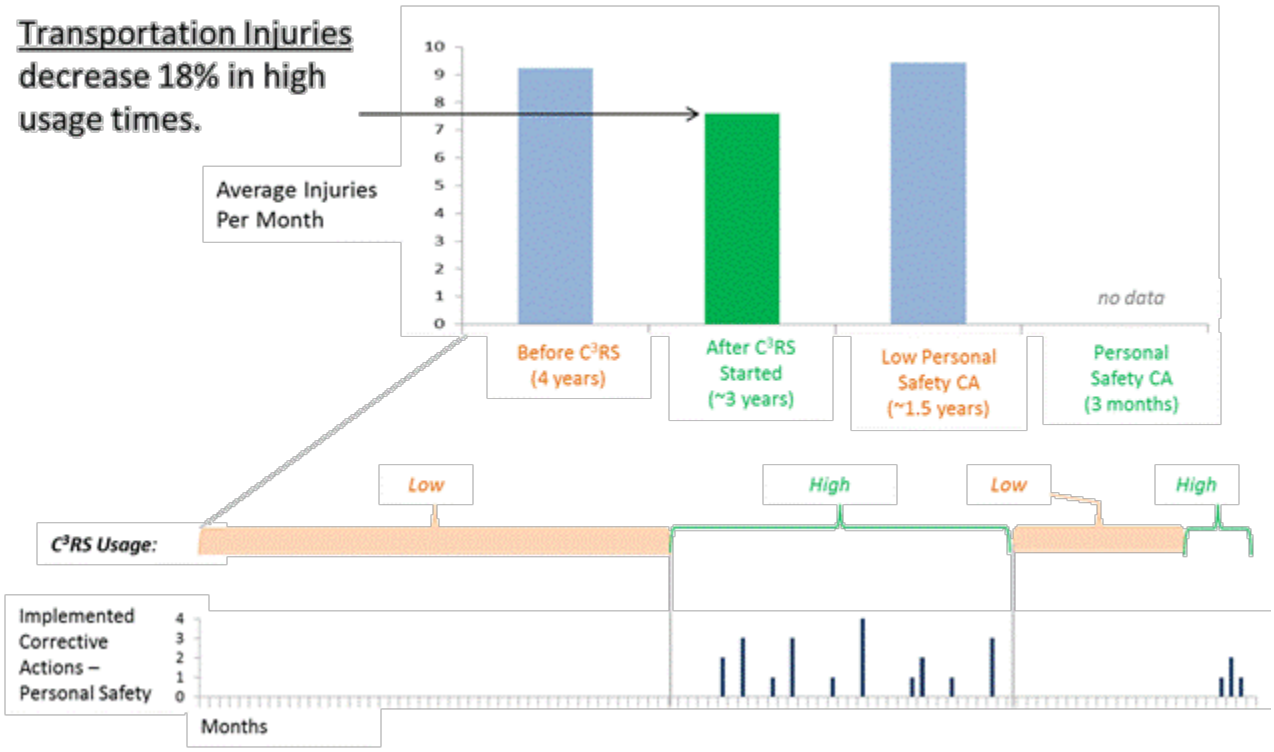


Figure 42. Transportation Injuries Decrease 18% at Site 2 during High Usage Times

Table 33. Site 2 Injury Results

Phase	Statistical Metric	C³RS Site Observed Value
Before and Low C³RS (4 years before C³RS, 1.3 years of low usage)	Average Number of Transportation Injuries per month	9.29
High C³RS Usage (3 years)	Average Number of Transportation Injuries per month	7.60
	Percent Change	-18%
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**0.003**

5.5.3 Corporate Data Results at Site 3

5.5.3.1 Derailments at Site 3

The Evaluation Team also examined derailments at Site 3. In the analysis, they included incidents with costs above the FRA reporting threshold and smaller non-reportable incidents. At this site, the team included both human factors and non-human factors derailments in the yard, as the interviewees indicated that the sample size of derailments was not large. Corrective action data showed that Site 3 implemented some corrective actions that were intended to prevent derailments, although not as many as Site 1. The Evaluation Team examined the derailment data and found that Site 3's derailments per cars moved non-significantly decreased by 20% when comparing the time before C³RS to the time after C³RS started (Table 34, Figure 43) ($F(276,320) = 1.24$, $p=0.03$, using an F-test for the ratio of two Poisson occurrence rates, Nelson, 2003). As this is a one-tailed p-value, the goal for "statistical significance" was 0.025. The result was considered to be not statistically significant; however, the Evaluation Team considered it worth reporting given the decreases in derailments at Site 1 and 2.

Table 34. Site 3 Derailment Results

Phase	Statistical Metric	C³RS Site Observed Value
Before C³RS (4.2 years before C³RS started)	Average Number of Yard Derailments per month (Human Factors and Non-Human Factors)	3.2
After C³RS (4.5 years)	Average Number of Yard derailments per month (Human Factors and Non-Human Factors)	2.6
	Percent Change	-20%
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**0.030**

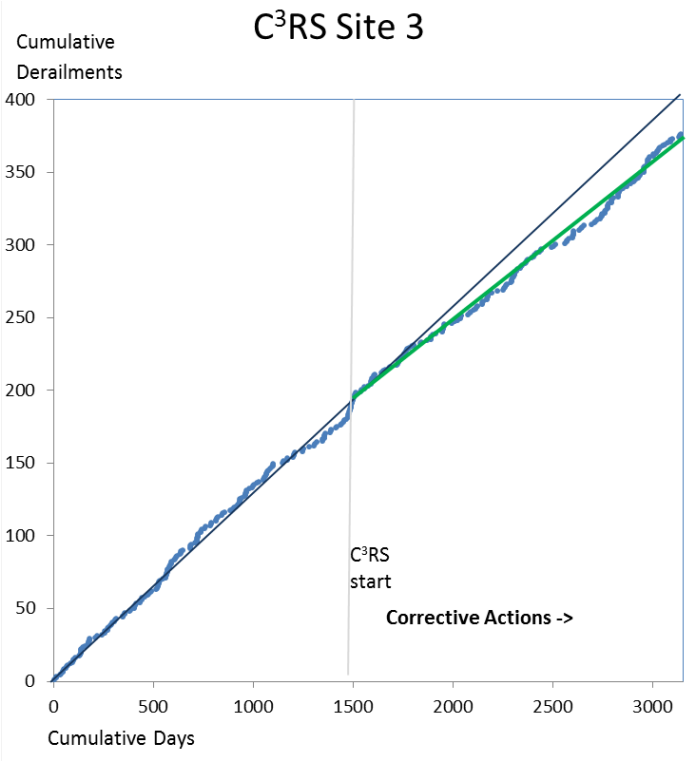


Figure 43. Cumulative Derailments at Site 3[37]

[37] An explanation for how to read cumulative charts was in Section 4.9.5.

5.5.3.2 Discipline Hearings at Site 3

C^{3}RS interviewees across sites talked about decreases in the number of discipline cases and the potential for cost savings. For every avoided discipline hearing, a railroad can save money that is usually spent on the investigation, preparation for the hearing, and associated travel costs. At Site 3, interviewees estimated that each hearing costs on average $10,000. Site 3 provided the Evaluation Team with data on the number of discipline hearings for operating rule violations. Site 3's discipline hearings significantly decreased by 39% when comparing the time before C^{3}RS to the time after C^{3}RS started (Table 35) ($F(274,316) = 1.64$, $p<0.001$, using an F-test for the ratio of two Poisson occurrence rates, Nelson, 2003).

Over the duration of the evaluation, Site 3 avoided 89 hearings, avoiding an estimated $890,000 in costs. The number of hearings avoided was estimated by multiplying the baseline average rate per year (see Table 35: approximately 39.5 per year) times the number of years in the period after C^{3}RS (5.7), then subtracting the number of hearings (136) actually held after C^{3}RS.

Table 35. Site 3 Discipline Hearing Results

Phase	Statistical Metric	C^{3}RS Site Observed Value
Before C^{3}RS (4 years before C^{3}RS started)	Average Number of Discipline hearings per month	3.3
After C^{3}RS (5.7 years)	Average Number of Discipline hearings per month	2.0
	Percent Change	-39%
	p-value (one-tailed, Poisson rate comparison (Nelson, 2003))	**<0.001**

5.5.4 Corporate Data Results at Site 4

5.5.4.1 Derailments at Site 4

Review of corporate safety data showed no impact on human factors derailments in the areas of Site 4 covered by C^{3}RS compared to non-C^{3}RS areas. This was consistent with the interviewees' assertion that the program did not have an impact on derailments.

Site 4's rate of derailments (human factors and non-human factors) significantly decreased 32% from before to after C^{3}RS corrective actions started (from 22 to 15 derailments per year, $F(150,186) = 1.44$, $p<0.01$). The comparison site also significantly decreased 40% (from 391 to

234 derailments per year, $F(2326,3320) = 1.69$, $p<0.01$). Given that the comparison site also decreased derailments, no convincing argument existed that C^{3}RS had any impact.

Site 4's rate of human factors derailments non-statistically significantly decreased 41% from before to after C^{3}RS corrective actions started (from 3.8 to 2.2 derailments per year, $F(24,32) = 1.56$, $p=0.12$). The comparison site significantly decreased 39% (from 103 to 62 derailments per years, $F(622,876) = 1.64$, $p<0.001$). This also confirmed that C^{3}RS did not have any impact on derailments at Site 4, where normalization data on cars moved was not provided to the Evaluation Team to allow for further analysis.

5.5.5 Summary of Corporate Data Results

Sites 1, 2, and 3 achieved positive impacts on safety as summarized in Figure 44. These three all reduced derailments between 20 and 41 percent. (Sites 1 and 2 had statistically significant reductions, and Site 3's reduction was not statistically significant.) Site 1 reduced their overall human factors incident cost by 53%, which was a large change but not statistically significant. Site 2 significantly reduced transportation injuries by 18%. Site 3 significantly reduced their discipline hearing frequency by 39%, resulting in an estimated cost savings of $890,000.

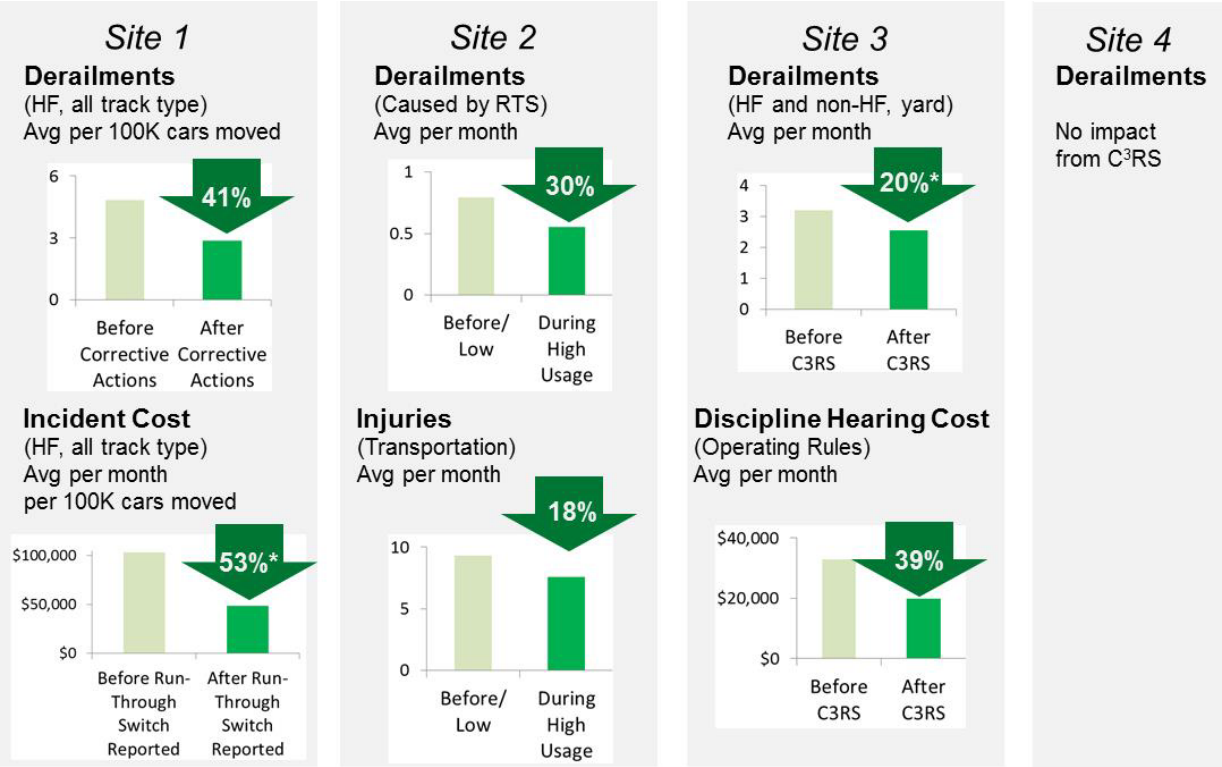


Figure 44. Summary of Safety Impacts[38]

[38] A decrease marked with an * indicates that it is not statistically significant.

6. Findings and Discussion

6.1 Summary of Findings

The C^{3}RS Evaluation Team assessed multiple types of data across the four demonstration pilots to answer three evaluation questions about implementation, impact, and sustainability. See Table 5 for a list of the data collected and summarized in the Results Section. Table 36 lists the evaluation questions and summaries of their answers. Discussions of the answers to the questions, which are the findings, are included in the sections below along with supporting results.

Table 36. Evaluation Questions and Answers

Evaluation	Question	Answer
Implementation	What conditions are necessary to implement C^{3}RS as planned in a demonstration?	Implementing C^{3}RS as planned was possible within transportation departments in the four C^{3}RS demonstrations. Three of the four sites performed all activities as planned from Figure 3, while the fourth site performed all the activities as planned except for one which it performed partially (implementing corrective actions). All four sites performed a sufficient amount of the activities as planned to justify evaluations of their impacts. Low reporting rates did not indicate poor implementation. The demonstration site with the highest rate of monthly C^{3}RS reports had the "poorest" implementation, did not show any bottom-line impact, and did not sustain its program post the evaluation period. It appeared that some reporters at this site were concerned more with indemnity and less about providing the details of the close call to prevent similar situations in the future. The common implementation factors for the three sites with the most favorable ratings were: support for C^{3}RS from a systemwide champion; labor enthusiastically involved in helping implement corrective actions; cooperation to achieve corrective action implementation; constructive dispute resolution; and shared positive perception of C^{3}RS's impact. The common opportunities for improvement for the four sites included: communication within and outside the program; PRT data analysis; PRT process efficiency; and tracking corrective actions.

Evaluation	Question	Answer
Impact	What is the impact of C^{3}RS on safety and safety culture?	C^{3}RS has demonstrated bottom-line impacts in three of the four sites in derailments, injuries, discipline hearings, equipment costs, and/or safety culture. Three of the four sites saw reductions in their derailments suggesting C^{3}RS shows promise in reducing safety risks in this area. The three demonstrations which received mostly favorable ratings on the implementation factors had significant bottom-line impact. The fourth site, with the most unfavorable implementation ratings, had no impact on safety. Labor perceived that safety culture between supervisors and labor improved in the presence of C^{3}RS-enabled communication. The sites with impact on safety also perceived improvements in safety culture scales related to the organization/management.
Sustainability	What factors help to sustain C^{3}RS long-term, beyond the demonstration?	Three of the four sites had favorable implementation ratings and observed impact. In two of those sites, there was agreement among the carriers' labor and management, the FRA, and national labor as to how C^{3}RS should continue to be organized and how it should function. Sustainment was achieved in those two sites. In the third site, where agreement could not be reached; sustainment was not achieved. The two sites that sustained were passenger railroads, and the two that withdrew were freight railroads. FRA has taken steps to support sustainability in the railroad industry. Those steps included continued funding for NASA, the allocation of dedicated staff, and the transfer of C^{3}RS from RD&T to the RRS. As a result of these efforts, C^{3}RS now has eight carriers participating (two original demonstration sites and six new carriers). Also the number of labor unions involved has grown from two in the original pilots (BLET and SMART-TD) to fifteen. (See Table 26 for a list of participating unions.) C^{3}RS began with just transportation and expanded to include engineering and mechanical crafts. Despite these steps from FRA, only railroads with passenger service are currently involved in FRA's C^{3}RS. It appears C^{3}RS is not as accepted by freight as by passenger carriers. This lack of acceptance applies to freight railroads of Classes 1, 2, and 3.

6.1.1 Implementation Findings

6.1.1.1 Implementation as Planned

Finding: Three of the four sites performed all activities as planned from Figure 3, justifying evaluations of their impacts. The fourth site did not perform as much for one of the activities as

planned (i.e., implementing corrective actions), but did perform it partially. Given the substantial effort that this fourth site made in the program and the fact that it did perform the other activities (on Figure 3), the Evaluation Team determined that lessons might be learned by conducting the impact analysis on this site too. Therefore, an impact evaluation on Site 4 was conducted to obtain lessons learned and for comparison purposes with the other sites.

Table 37 summarizes the results supporting this finding.

Table 37. Summary of Results on Implementation Completeness

Method	Result
C^3RS Reporting Data	Eligible workers at all four demonstration sites submitted C^3RS reports to the Third-Party. The Third-Party collected the reports, and once sanitized, provided them to PRTs at each of the four sites.
Interview and other Qualitative Data	Sites 1 through 3 were assessed as having implemented the program as planned because out of the six activities listed in Figure 3, they completed all of them. Site 4 only partially implemented "corrective actions implementation" as planned, because labor did not participate in the actions in which they had control and could participate.
Corrective Action Data	All four sites implemented corrective actions based on the data they received from C^3RS and the analysis the PRT performed. Site 4 did not implement a lot of actions but did some.

Sites 1, 2, and 3 implemented activities as planned up to the implementation of corrective actions (Figure 3). Reporting data showed that eligible employees submitted hundreds of reports, which the Third-Party processed and sent to PRTs. Interview data showed that the labor, management, and FRA participated on the PRTs at each site. The level of participation and effectiveness varied among sites, with some showing stronger support from labor and management than others. Senior management, sometimes in the form of a Support Team, reviewed and implemented corrective actions. Corrective action data, while not consistently organized, demonstrated that they implemented numerous corrective actions, thus justifying an evaluation of impacts.

Site 4 also received reports, analyzed them on the PRT, and made recommendations. However, they had less activity on corrective action implementation. Local labor was not involved in the implementation, leaving it to management. Nonetheless, some actions were implemented, which justified evaluating their impact.

The two areas of Figure 3 with less activity for all four sites were: "monitors corrective actions" and "railroad shares information." The four sites did minimal monitoring of corrective actions for implementation progress and impact. Interviewees indicated that there was insufficient information sharing after the railroad implemented corrective actions. When sites tried to share information, they used newsletters, signs, and training to let employees know what C^3RS was accomplishing. However, some interviewees from outside the PRT could not cite examples of corrective actions.

6.1.1.2 Low Reporting Rates do not Indicate Poor Implementation

Finding: The demonstration site with the highest rate of monthly C^3RS reports had the "poorest" implementation, did not show any bottom-line impact, and did not sustain its program post the

evaluation period. It appeared that some reporters at this site were concerned more with indemnity and less about providing the details of the close call to prevent similar situations in the future. In contrast, the three sites with fewer reports submitted received more "good" ratings for implementation and achieved bottom-line impacts.

Table 38 summarizes the results supporting this finding.

Table 38. Summary of Results on Low Reporting Rates may not Indicate Poor Implementation

Method	Result
C^3RS Reporting Data	Site 4's number of reports per 100 eligible reporters was more than three times as many as any of the other three sites.
Interview and other Qualitative Data	Ratings of implementation factors at Site 4 were either "poor" or "fair" across six of the eight main implementation factors: carrier responsibility, labor responsibility, shared responsibility, ability to implement corrective actions, effective dispute resolution, and perceived value. Site 4 decided to withdraw from C^3RS at the end of the demonstration pilot. Interviewees at Site 4 did not believe that C^3RS had a positive impact on safety or safety culture. In contrast, Sites 1, 2, and 3 did not have any main implementation factors ratings of "poor."
Survey Data	Site 4 respondents also became less willing to report to C^3RS by the time of the final survey. In contrast at Site 1 and 2, respondents were more willing to report to C^3RS.

Content of reports is considered more important than the volume of reports. "Content" is defined as substantive detail providing information about the close call and the circumstances in which it occurs. Volume is the number of reports per month, normalized by the number of eligible employees.

Looking at the implementation findings, it is reasonable to assume that the high "volume" of reports with insufficient "content" and the subsequent lack of impact can be tied to other characteristics of Site 4's C^3RS implementation. Labor distrust of management's motives led to a "get out of jail free" mindset that employees had with respect to C^3RS reporting. Management at the site introduced a more coercive discipline system that sent conflicting messages to labor. Labor did not help with corrective actions, which affected the number of corrective actions implemented and their effectiveness. They lost senior sponsorship. Also, the initial Support Team did not adequately respond to C^3RS. Thus, activity needed for effective corrective actions was delayed or missed. Disputes resulted in discipline, which further caused employees to be distrustful and not provide details in their C^3RS reports. An attempt to change the IMOU failed. As a result of these issues, Site 4 did not have enough detailed reports to understand root causes and to devise and implement impactful corrective actions.

When a C^3RS program is implemented, the emphasis should be placed on the purpose of the reports as a way to share information. While the information about discipline avoidance can be included as a benefit for labor, it is not the primary purpose of C^3RS. There can be value in including discipline, considering all the previously unknown information other sites uncovered.

6.1.1.3 Common Implementation Factors Associated with the Most Favorable Implementation

Finding: Three sites had five or more "good" ratings on the main implementation factors. The common implementation factors for those three sites with the most favorable ratings were: support for C^{3}RS from a system wide champion; labor enthusiastically involved in helping implement corrective actions; cooperation to achieve corrective action implementation; constructive dispute resolution; and shared positive perception of C^{3}RS's impact.

Table 39 summarizes the results supporting this finding.

Table 39. Summary of Results on Implementation Factors Associated with Impact

Method	Result

Interview and other Qualitative Data	Analysis of implementation revealed that the three sites with at least five out of eight implementation factors having ratings of "good" or "very good" had some common characteristics (Figure 30). They included:
	• Carrier Responsibilities: rated "good" or "very good" on at least four of their five detailed factors
	• Labor Responsibilities: rated as "good" or "very good" on at least two of their four detailed factors
	• Stakeholders cooperating together:
	  ○ Negotiated IMOU
	  ○ Supported training
	  ○ Cooperated on corrective actions
	  ○ Resolved disputes in a constructive way that encouraged continued future reporting
	• All stakeholders agreed that C^{3}RS was valuable

The implementation analysis of the interview data revealed many commonalities among the three sites with good implementation (i.e., Site 1, Site 2, and Site 3) (Figure 30). A system wide champion offered support for C^{3}RS. Labor enthusiastically helped implement corrective actions. Cooperation helped achieve corrective action implementation. When disputes over discipline arose, they resolved them, and stakeholders moved on. Stakeholders had a shared positive perception of C^{3}RS's impact on safety culture, safety, corrective actions, and cost savings.

All three sites had an effective system wide champion at the senior level of their companies. Local middle management support varied over the duration of the pilots, but overall remained adequate (each site had "good" ratings for either the managers on the PRT or from local management sponsors). Management also provided sufficient resources for PRT operations.

The three sites had "good" or "very good" ratings for labor helping implement corrective actions, which appeared to be essential. For example, if a corrective action involved training or communication, labor PRT members went into the field and/or classroom to complete it. The three sites were also good at promoting C^{3}RS reporting and/or serving on the PRT. This involved talking about C^{3}RS informally among their coworkers and telling coworkers that it was

important to report close calls for reasons other than discipline protection. This encouraged the workforce to provide a sufficient flow of reports enough for the PRTs to develop corrective actions to affect safety. Future C^{3}RS implementations need to pay special attention to ensure that these factors are being implemented thoroughly.

The sites with favorable implementation also all had "good" ratings for cooperation among stakeholders to implement corrective actions. This led them to implement enough corrective actions to all have a "good" perceived value of C^{3}RS.

The site with the least favorable implementation had "poor" ratings for dispute resolution (e.g., about such issues as whether reports fall within the scope of the IMOU discipline immunity), while the other three had higher ratings. This showed that sites needed to resolve disputes in a way that maintained eligible reporters' trust. If employees lose trust in the system, they cannot be motivated to provide reports with enough detail to determine contributing causes and justify the costs to run a C^{3}RS program. For example, at Site 4, a C^{3}RS reporter was disciplined for an offense related to his C^{3}RS report and lost six months off work. The three sites with favorable implementation were able to move on after disputes in a way that generated trust in the intention or spirit of the program and received "good" ratings. These resolutions were perceived as supporting the confidentiality of the reporters and promoted the submission of reports that were complete enough to help in the determination of root causes to enable the development of meaningful corrective actions. Credible corrective actions helped to justify the costs to run a C^{3}RS program.

These findings speak to the tactics needed for good implementation. First, it was not necessary for all implementation factors to operate at their highest possible level. As is evident in Figure 29, the three sites with overall good implementation did not have a large number of "very good" ratings. Rather, a few such ratings and a much greater number of "good" ratings for each stakeholder were sufficient. Second, a different combination of factors that operate reasonably well can lead to successful implementation. The three sites with good implementation had some variation between them on which detailed factors were rated as "good" (Figure 30). It may be that implementation success depends on the number of factors that operate well, rather than the specific ones that do. The exception to these observations is the role played by the FRA, where "very good" ratings were observed for both "funding" and "IMOU."

6.1.1.4 Implementation and Impact

Finding: Sites with good implementation demonstrated positive impacts.

Table 40 summarizes the results supporting this finding.

Table 40. Summary of Results on Implementation and Impact

Method	Result
Interview and other Qualitative Data	Analysis of implementation factors and impact data revealed that the three sites with five or more "good" or "very good" favorable ratings on the main implementation factors had a positive and significant impact on safety and other corporate metrics.
Corporate Archival Data	At Site 2, the data revealed that during times of "low usage" of C^{3}RS, there was less impact than during times of "high usage."

The implementation strength analysis of the interview data revealed that Sites 1, 2, and 3 had "good" implementation on more than half of their main implementation factors, and they had an impact on safety metrics. At those three sites, many detailed factors received ratings of "good" (22–26 out of 35 detailed factors were "good" or "very good" and 0–2 were "poor"). In contrast, Site 4 had the poorest implementation (12 "good" or "very good," 11 fair, and 12 "poor" on their detailed factors) and showed no significant improvements in safety impacts. See Section 6.1.2 for more findings on the details of impact.

At Site 2, during times of "low usage" of C^{3}RS, the positive impact from C^{3}RS disappeared (Section 5.5.2). Derailments caused by run-through switches reverted to similar rates as before C^{3}RS during time periods when eligible employees did not report run-through switches to C^{3}RS. Transportation injuries reverted to similar rates as before C^{3}RS was implemented. This indicates the need to continue "good" implementation for continued impact.

6.1.1.5 Opportunities to Improve

Finding: The common opportunities for improvement for the four sites included: communication within and outside the program; PRT data analysis; PRT process efficiency; and tracking corrective actions.

Table 41 summarizes the results supporting this finding.

Table 41. Summary of Results on Similar Improvements for Implementation

Method	Result
Interview and other Qualitative Data	Most sites mentioned common aspects to their programs that needed to be improved. At least 3 of the 4 sites had "fair" or "poor" ratings in these areas: communication within and outside the program, PRT data analysis, PRT process efficiency, and tracking corrective actions (Figure 30).

C^{3}RS would be more powerful and useful if these activities were improved, but the program can make a contribution to safety even without a high level of achievement for these activities. The activities include communication within and outside the C^{3}RS program, PRT data analysis, process efficiency, and the tracking of corrective actions.

Two types of communication were mentioned by interviewees as needing improvement: 1) between Support Teams and PRTs and 2) between the C^{3}RS program's PRT and Support Team and labor and management in the rest of the carrier.

The communication between the PRT and Support Team tended to be "fair" at most sites. Support Team members expressed frustration that corrective action recommendations did not include information on the frequency and magnitude of the safety problem, making it difficult to prioritize resource allocation across them. The PRTs did not provide the business case (benefits and costs) for their recommendations. On the other hand, PRT members expressed frustration that the Support Team did not communicate feedback about issues considered in making decisions about which corrective actions to approve and the status of corrective actions (e.g., approved, not approved, implemented, or delayed). The C^{3}RS Implementation Team did not explicitly define who should take on the role of providing a business case for corrective actions. This ambiguity left opportunity for responsibilities to fall between the cracks and for each group to blame the other for not taking on this work. The PRT also did not have the skills to provide

the Support Team with the business case that they wanted and the Support Team did not communicate their expectations or provide training on how to do this.

The communication from the C^{3}RS Teams (i.e., PRT and Support Team) to others at the carrier tended to be "fair." Workers and managers outside the program often did not know much about C^{3}RS accomplishments. Eligible reporters stated that without feedback on corrective actions implemented as a result of their reporting, they were not sure their reporting mattered. Feedback on accomplishments motivated them to provide more detailed reports. Managers also needed to understand the accomplishments occurring from the reporting which would then help encourage their employees to report. Senior managers needed to know about accomplishments to justify continued investment. Some of the lack of communication was caused from concerns about confidentiality. Some of it was due to lack of time/resources to create more communications, like newsletters.

PRT data analysis was considered "fair" at three of the four sites. As taught, PRTs analyzed one case at a time using the MCIA tool to identify root and contributing causes and developed corrective actions for the identified contributing causes. Over time, the PRTs gained experience and access to a continually mounting source of close call data. As a result, they began to work on groups of related cases. However, their efforts stopped short of providing a deeper understanding of the contributing causes and related potential corrective actions, because the PRT did not collect and analyze additional data on their own (through surveys or observations or directing reporters to report on certain conditions) and/or obtain data from other parts of the carrier such as safety data. Therefore, their corrective actions were limited to the information reporters were submitting. Without secondary and corroborating sources of information their data was not as robust as it could have been. They could not compare different corrective actions and help management to prioritize their recommendations as they made decisions about how to spend limited resources. The PRTs were not staffed adequately in this regard. They could have benefited from a specialist in an area, such as continuous process improvement or operations research. At one site, when staffed with a problem-solving expert, the contributing factors and corrective actions became more robust.

Opportunities to improve PRT efficiency also existed, mostly with respect to the frequency and makeup of PRT meetings. The original concept of operations of C^{3}RS, as negotiated by each PRT, had all members of the PRT attending all meetings. This was inefficient, especially for large PRTs, because as deliberations proceeded, the PRT scribes needed to manually enter data, thereby making the rest of the PRT wait. As the workload increased, sub-teams met in-between whole group meetings to perform the data entry, complete administrative tasks, group similar reports, and prepare reports for productive discussion.

Despite this improvement, the C^{3}RS process could not escape a major inefficiency: the expectation that every case had to be reviewed in detail in person by the whole PRT. This requirement made sense when C^{3}RS first began, because the program developers wanted a system that would be perceived by the workforce as being responsive to their observations about safety. Bundling cases into related groups helped with this problem, but not enough to solve the efficiency issues.

This problem also became an issue for some Support Teams who had limits with respect to the number of corrective actions they could consider. Similarly, railroads had limitations in the

number of changes they could manage or afford to implement at any given time. Support Team members had many other responsibilities and spent limited time working on C^{3}RS.

Finally, an opportunity existed to improve the tracking of corrective actions. Tracking refers to recording and monitoring the status and impact of corrective actions. The PRTs sometimes found a local corrective action that could be coordinated with a manager they knew and implemented quickly. This provided an efficient way to make things happen. However, as PRT members rotated, and with incomplete documentation of the actions implemented, the knowledge was lost and could not inform future contributing cause analyses or corrective action recommendations. Moreover, due to the informal nature of this immediate implementation with a local manager, new procedures might be tried and found successful (or not), but they were not always institutionalized by revising the procedural manual or policies. FRA is planning to add a more robust tracking feature to the MCIA tool to help address this opportunity.

Another aspect of tracking a corrective action is tracking its impact. The PRT struggled to know if the corrective actions were accomplishing their goals. The railroads collected metrics, but the PRT did not always have access to that data. During the demonstration, the Evaluation Team helped mitigate this issue, because they provided impact assessments to the sites. After the evaluation, all sites must assess impact for themselves.

6.1.2 Impact Findings

6.1.2.1 Impact on Derailments

Finding: Derailments decreased at three sites in the presence of C^{3}RS: 1) corrective actions and 2) communication enabled by discipline protection. This was the most common impact in the Evaluation suggesting that C^{3}RS shows promise in improving safety in this area.

Table 42 summarizes the results supporting this finding.

Table 42. Summary of Results on Derailments

Method	Result
C^{3}RS Reporting Data	There were many C^{3}RS reports on derailments and run-through switches, which can cause derailments. Switching and derailment together made up 26% of the C^{3}RS-analyzed-cases.
Corrective Action Data	Sites implemented corrective actions to reduce derailments and run-through switches.
Interview and other Qualitative Data	Interviewees explained that there were two ways that derailments could decrease in the presence of C^{3}RS: 1. Reports on derailments and switches led to corrective actions which addressed the causes of derailments. 2. When an employee reported on a run-through switch, they received discipline protection and were expected to call their supervisor. Then the switch was repaired immediately, preventing a future train from having a derailment due to a damaged switch.

Method	Result
Corporate Archival Data	Derailment rates were reduced at three sites, all three by an operationally substantial amount: • At Site 1, Human Factors derailments significantly decreased 41% when comparing a multi-year baseline period to the time after C^{3}RS corrective actions began. A comparison site with similar characteristics in the same railroad was also checked and did not change. The traffic volume had a lot of variation during the evaluation period, so the rates were normalized by cars moved. • At Site 2, there was significant 30% reduction in the types of derailments caused by run-through switches during times of high usage. Site 2 had times when run-through switches were not being reported, and during those times derailments were more likely to occur. Greater snowfall during the C^{3}RS period seemed to be related, so it was used as a covariate when checking for statistical significance. • At Site 3, there was a 20% reduction in derailments. The statistical test showed significance just outside the limit, but was considered to be "heading in the right direction." Site 3 had less reporting and corrective actions related to derailments and switching than Sites 1 and 2, so this result was not surprising.

Reports on either derailments or switching represented 26% of the cases analyzed by the PRTs. Because switching problems are related to derailments, this provided a great deal of information for the PRTs to learn about the root causes of derailments. The two sites with the biggest impacts on derailments had a third of their cases dealing with either derailments or switching. The site with the biggest impact on derailments allowed discipline protection for minor derailments that actually occurred under the FRA reporting threshold, so they learned the most about contributing causes.

The three sites implemented many corrective actions to reduce derailments and switch damage, such as improved communications in the yard, adding visual checks to yard processes before movement, training on backing up, a new policy about derail removal, improved switch visibility, and a poster about commonly run-through switches.

The evaluation discovered another way to reduce derailments, besides corrective actions, which was particularly apparent at Site 2. Interviewees said that, in the past, when a crew had ran-through a switch, they would often try to fix it themselves to avoid being disciplined. These repairs were often not adequate, because the people making the repairs were not fully qualified or trained to repair switches. As a result of the incomplete repairs, subsequent trips through the switch could result in a derailment. With the discipline afforded by C^{3}RS, the crews felt comfortable telling their supervisor about the damaged switch, allowing for repairs to be made by qualified mechanical personnel. When Site 2 was reporting on run-through-switches, their related derailments decreased. When Site 2 was not reporting on run-through-switches, then their related derailment rates increased back to similar rates as before C^{3}RS. This shows that impact can be achieved when a site works the program and can be lost when it stops actively working the program. C^{3}RS is not like a vaccine, where a single fix solves the problem. Instead it is like an exercise program that requires ongoing effort to produce sustained results.

This improvement at Site 2 caused previously unknown run-through switches to become known, increasing the number of recorded run-through switches. Interviewees indicated that this produced some negative perceptions of C^{3}RS. Efforts were needed to inform people that the situation was not worsening; it was just becoming known. As a result, Site 2 began further efforts to try to prevent run-through switches.

6.1.2.2 Impact on Injuries

Finding: Injuries decreased at one site in the presence of corrective actions such as improved job safety briefings. More focus on encouraging employees to submit close calls that related to injuries and implementing corrective actions to reduce injuries is needed to make more improvements in this area across sites.

Table 43 summarizes the results supporting this finding.

Table 43. Summary of Results on Injuries

Method	Result
C^{3}RS Corrective Action Data	Site 2 made multiple improvements to their job safety briefing, making attempts to increase the awareness of their crews. Other corrective actions were implemented as well. The Evaluation Team discussed the corrective actions with the PRT to determine which could reasonably be expected to have an impact on injuries. However, Site 2 did have a time period during the C^{3}RS demonstration where multiple external factors slowed down the usage of C^{3}RS and the implementation of corrective actions related to injury prevention.
Corporate Archival Data	Site 2's transportation injury rates decreased 18% when comparing multiple years before C^{3}RS to the time during high usage of C^{3}RS. During times of low usage, the injury rates returned to values similar to those before C^{3}RS, indicating that usage of C^{3}RS was related to impact. Managers at Site 1 did not believe that the corrective actions were related enough to injuries to justify providing the Evaluation Team with the data. Site 3 and 4's injury rates did not improve.

Site 2 reduced their injury rates by 18% when comparing "high usage" of C^{3}RS times (when they were implementing corrective actions that could attribute to injuries) to "low usage" times (either before C^{3}RS or when injury-related corrective actions were not being implemented). This pattern indicates more support for the finding that C^{3}RS can reduce injuries. It also confirms the finding that impact can be achieved when a site works the program and can be lost when it stops actively working the program.

In contrast to derailments, which C^{3}RS did reduce, three of the four demonstration sites did not have a measureable impact on injuries. This result makes sense because few C^{3}RS reports dealt directly with injuries, highlighting an opportunity for C^{3}RS to increase its impact in the future. If employees had been encouraged to submit more injury-related close calls to C^{3}RS, it is reasonable to assume that injury prevention corrective actions would be developed and implemented, thereby reducing injuries.

6.1.2.3 Impact on Discipline

Finding: Discipline hearings and associated costs decreased in the presence of C^3RS protection, thus helping to offset the cost of the C^3RS program.

Table 44 summarizes the results supporting this finding.

Table 44. Summary of Results on Discipline Hearings

Method	Result
Interview and other Qualitative Data	Interviewees at Site 1, who were involved in discipline hearings, cited a 90% decrease in the number of hearings. The reduction was attributed to discipline protection from the C^3RS IMOU and FRA waivers. This was to be expected; less discipline = less hearings. Interviewees at Site 3 estimated the cost savings associated with the reduction in their hearings to be approximately $10,000 per hearing in labor and travel.
Corporate Archival Data	Site 3 reduced the number of discipline hearings by 39%. The reduction means that they avoided 89 hearings over the course of their demonstration, avoiding an associated $890,000.

At multiple sites, interviewees confirmed that one of the benefits of C^3RS was avoiding the paperwork, time, and effort of discipline investigations. Site 3 interviewees provided estimates of the savings and said that a single discipline investigation with a hearing costs approximately $10,000. The cost came from the time off work for the employee being investigated, drug and alcohol testing, the time the managers spent performing the investigation, and the time spent on the hearing itself.

Two sites provided data to quantify the reductions. At Site 1, interviewees reviewed their records which showed an approximately 90% reduction in hearings, which presumably would have translated into a 90% reduction in costs. At Site 3, carrier records indicated that over the course of the demonstration approximately $890K was avoided due to the reduction in hearings (39% lower annual rates of hearings as compared to previous years). These cost savings helped justify the expense of the program and confirm the expectation that carrier resources were shifting from discipline to prevention.

6.1.2.4 Impact on Safety Culture

Finding: Labor perceived that safety culture between supervisors and labor improved in the presence of C^3RS-enabled communication. The sites with an impact on safety also perceived improvements in safety culture scales related to the organization/management.

Table 45 summarizes the results supporting this finding.

Table 45. Summary of Results on Safety Culture

Method	Result
Survey	All the sites, regardless of how good their implementation was, had increases in the safety scales related to supervisors, e.g., Supervisor Fairness (SF), Supervisor-Employee Relationships (SER), and Raising Concerns with Supervisors (RCS). However, only the sites with an impact on safety had any improvements in safety culture related to the organization/managers (Sites 1, 2, and 3). Two sites had decreases in a coworker related scale (Sites 1 and 4).
Interview and Other Qualitative Data	Interviewees at Sites 1, 2, and 3 had many comments about how C^3RS changed the culture previously focused on blame and discipline to a culture of communication and problem solving. For example: "Once a manager gets involved in C^3RS, it changes their thinking. They are ready to find a solution." "My workers feel more comfortable approaching me." In contrast, Site 4 interviewees talked about how multiple disputes and discipline of some reporters led to reduced trust in C^3RS.

The **Supervisor** railroad safety culture scores showed that labor observed differences in the fairness of supervisors, supervisor-employee relationships, and their ability to raise concerns with supervisors. These differences were measured between the time that C^3RS was first starting to the end of their 5-year demonstration. This improvement in safety culture held true for all three Supervisor scales at all three sites that took the survey (i.e., Sites 1, 2, and 4). The only exception was Site 2, where the final average in Raising Concerns with Supervisors appeared higher but did not significantly improve. Site 3, using their own survey, also had improved views of supervisors.

Interviews at Sites 1, 2, and 4 corroborated these survey results, and they discussed how C^3RS opened up communication and comfortableness between employees and their supervisors. Workers became more comfortable approaching their managers to talk about safety observations. Managers began thinking more about collaboration with workers to find solutions to problems. Not every manager fully embraced this new culture; some remained skeptical and focused on discipline, but a cultural shift began. This explains why Supervisor scales were the most likely to improve.

For the **Organizational/Managerial** scales, Site 1 improved both Labor-Management Relationships and Management Safety. Site 2 had a small improvement in views of labor-management relationships. None of the sites surveyed by the Evaluation Team showed improvements in Organizational Concern for Employees. Site 3, using their own survey, improved their views of Organizational Concern for Employees as well as other Managerial scales. Sites 1 and 3 had two different safety programs, so perhaps in conjunction, employees felt better about management. Only the three sites that had improved safety also had improvement on any Organizational/Managerial scales, in contrast to Site 4 that improved neither. This may be because labor attributed improved safety as an indicator of the organizations' commitment to their safety.

The **Coworker** scales revealed a lack of consistent change in views of coworkers, with increases, decreases, and unchanged scales. Interview data did not provide much explanation for these

inconsistent findings. A plausible explanation is an artifact of measurement. These scale scores had very high values at baseline, thus restricting the range of possible change, and decreased the ability of statistical analysis to detect true effects. Site 2 did better than Site 1 on the coworker scales; however, the reasons are not clear, and Site 1's views were still high.

The main observation was that Site 4 had a decreased view of helping behavior, in agreement with the interview results showing that labor did not help management implement corrective actions. Both sites that were not sustained after the demonstration (i.e., Site 1 and 4) had some decreases in views of coworkers, with the larger decrease being for Site 4. At Site 4, the decrease may have been caused by the issues with C^3RS implementation, for example labor's lack of involvement in corrective actions.

Across several Organizational/Managerial and Supervisor scales, Site 2 had less improvement than Site 1. This may be because Site 2 had other external events that may have dampened their perceptions, such as some unrelated tension in labor-management relations over contracts.

Site 4 had fewer scales improve than the other sites. Given that Site 4 had poorer implementation and had decided to withdraw from C^3RS immediately at the end of their demonstration, this is to be expected. What is surprising is that Site 4 had statistically significant improvements in the Supervisor safety scales, given that their interviewees were so pessimistic about there being any positive perceived value of C^3RS. Site 4's lack of improvement in any Organizational/Managerial scales and their decrease in Helping Behavior are most closely aligned with their implementation difficulties and lack of impact on safety from C^3RS.

6.1.3 Sustainability Findings

6.1.3.1 Sustainable Demonstration Sites

Finding: Good implementation, impact, and agreement on a revised IMOU amongst national labor, FRA, and internal labor and management stakeholders are needed for a site to be sustainable after the initial operating period. All four demonstration sites needed to revise their IMOUs after they completed the demonstration period. Two went smoothly leading to continuing the program, and two did not go smoothly leading to withdrawal from FRA's C^3RS. For the two sites that withdrew, the largest points of tension in revising the IMOUs concerned whole crew reporting or the use of NASA as the Third-Party. The agreement on how to address emergent IMOU issues seems to be a key differentiating factor for which sites sustain and which do not.

Table 46 summarizes the results supporting this finding.

Table 46. Summary of Results on Sustainable Sites

Method	Result
Interview and other Qualitative Data	Two of the four demonstration sites decided to continue and are still participating in FRA's C^{3}RS as of this writing. Demonstrations that sustained had six-seven "good" implementation main factors, with support and collaboration between carrier labor and management, FRA, and the national labor unions. Sustained sites had agreement among stakeholders concerning the Third-Party, sharing of data with industry, and reporter-known events. The two sites that left FRA's C^{3}RS could not come to agreement with the FRA, the national unions, and their own labor and management regarding the future IMOU.
Corporate Archival Data	Sustained sites had positive impact on safety impacts. (Although, so did Site 1, which later left, so impact on safety seems to be necessary but not sufficient for sustainability.)

Above, the characteristics of "good" implementation were discussed. When it comes to sustainability, both good implementation and proven impact were needed (Figure 29). The two sites that decided to continue with C^{3}RS long-term both had mostly "good" implementation ratings (25–26 of 35 detailed implementation factors and six-seven of eight main factors). The sites that continued also achieved a positive impact on safety and safety culture.

Implementation and impact by themselves were not enough to achieve sustainability (Figure 29). Sustainability required agreement between labor, management, and FRA on emergent IMOU issues. At Site 1, the changes between the demonstration and using NASA as the Third-Party were not satisfactory, because they were concerned with confidentiality. Site 1 was not comfortable with the concept of the national database, because they did not believe their site and employee anonymity would be protected. At the end of their demonstration, Site 1 did not anticipate this lack of consensus would occur, and the sustainability issues arose in later years. At Site 4, their emergent tensions with the IMOU conditions caused them to attempt unsuccessfully to revise it to encourage more complete reports. They wanted to require all members of the crew to report. National labor was not supportive of the proposal to require all crew members to report if they wanted discipline protection, even though local labor wanted the change for the purpose of collecting more complete close call reports. Unable to reach agreement on how to run C^{3}RS in the future, Sites 1 and 4 withdrew.

6.1.3.2 Sustainment in the Industry

Finding: FRA has taken significant steps to support long-term sustainability in the railroad industry and has expanded the program across railroads and unions. The expansion was almost entirely centered on passenger rail. Class 1 railroads and shortline freight railroads (except for one) were not involved as of the time of this report. This suggests that a large percentage of the industry is not being addressed by C^{3}RS, and this shortfall should be resolved.[39]

[39] The Association of American Railroads states that freight railroads had more than 169,000 employees. (Accessed online July 27, 2017).

Table 47 summarizes the results supporting this finding.

Table 47. Summary of Results on Sustainability in the Industry

Method	Result
Interview and other Qualitative Data	Data revealed that FRA used the positive C^{3}RS evaluation results to justify funding within the agency and from Congress. They also used the published results to motivate new sites to join. FRA continued to fund the Third-Party and set up the Human Performance Division to support C^{3}RS. Eight carriers and many national labor unions supported C^{3}RS programs as of the writing of this report (Figure 31). All of the eight sites involved in C^{3}RS as of the writing of this report were passenger carriers (two of the demonstration sites plus six new sites). No Class 1 railroads were involved, and with the exception of some freight business in one shortline carrier, no Class 2 or 3 railroads have joined. Of the four demonstration pilot sites, the two freight carriers eventually withdrew, while the two passenger carriers remained in the program long-term post demonstration.

Long-term sustainability was observed as FRA continued to recruit new railroads to participate, and supported C^{3}RS with funds and staff. The FRA C^{3}RS members used the evaluation results to gain support from FRA to set up the HPD. The evaluation impact results were also used in presentations and publications to justify C^{3}RS to potential new sites.

Two of the demonstration sites, six new carriers, and many additional labor unions covering transportation, engineering, and mechanical departments were participating in C^{3}RS as of this report (Table 26). Many of them operated passenger trains on the Northeast Corridor. As of this report, no Class 1, 2, or 3 freight carriers were participating in C^{3}RS. (The only exception is the freight part of the Strasburg Railroad, who also carried passengers.) Lack of interest by freight carriers was evident both in the demonstration sites and in the FRA's ongoing recruitment efforts. Of the four demonstrations, the two passenger carriers remained with C^{3}RS long-term, while the two freight carriers did not. This pattern represents a challenge that the FRA should address if it wants C^{3}RS to help railroads across all sectors of the railroad industry. In particular, the fact that one freight demonstration site was still considering proactive safety programs shows that close call reporting can be appealing to a large freight railroad, even if FRA's particular, industry-wide model of C^{3}RS is not. The reasons for this lack of appeal across all the Class 1's and shortlines are unknown.

The lack of Class 1 railroads was a challenge to FRA, because they wanted more sites to participate in the national model, so that the de-identified database could be populated and used to issue safety alerts to the industry. To address that challenge, the railroads' interest in C^{3}RS should be seen in system terms wherein C^{3}RS is embedded in a government/industry/social/economic/political setting that makes the program more appealing to some types of railroads than to others. A framework for understanding the differential appeal includes both the legacy of C^{3}RS's origin and conditions affecting it now.

C^{3}RS's starting conditions had an effect on its development. At a particular time in its policy making trajectory, the FRA found it worthwhile to probe the industry on the possibility of a close call reporting system through the close calls workshop in 2003. Also at that same time, a

particular set of conditions were at play in the railroad industry, such as a history of restricted communication between labor and management about safety and a desire to reduce incidents. Those conditions also included the industry's relationship with FRA, relationships among the stakeholders in the industry, each stakeholder's perception of its self-interest, and the business climate. As a consequence, enough enthusiasm existed in government, industry, and labor to recruit four sites into the 5-year demonstration program. Considering the differences between the conditions then and now may help FRA to understand more about railroads' current level of interest in starting C^3RS.

However, understanding those initial conditions is not sufficient because over time other factors may have come into play affecting the context in which the stakeholders are operating, and their reactions to it. From a mission, policy, funding, and accountability point of view, FRA was satisfied enough with C^3RS to commit to continuing its efforts, based on the positive impacts shown through the demonstration and the enthusiasm of the eight involved railroads. FRA committed to a particular configuration of C^3RS. As with the four original sites, FRA should explore what conditions are at play in the industry to incentivize railroads to participate. In particular, why is the program more attractive to passenger than freight railroads? How can freight needs be addressed? In addition, FRA should continue to monitor and explore changing conditions in the industry and emerging concerns of both labor and management, possibly making revisions to the IMOU. As an example, the impact of new surveillance technology, such as cameras in the locomotive cab, needs to be determined.

6.1.4 Summary of Findings from the Demonstration

The C^3RS demonstration revealed that implementing a close call reporting system is possible within the U.S. railroad industry. In order to achieve impacts, favorable implementation is needed, including visible carrier and labor commitment to the program and cooperative sharing of responsibilities among carrier labor and management, national labor, and FRA. If implementation is favorably executed, then C^3RS can lead to bottom-line impacts on safety and safety culture. The submission of higher than average numbers of C^3RS reports by employees may not guarantee success and may even be a warning sign that reporting is being used primarily for discipline avoidance.

If the intention of the program (i.e., providing information on close calls to help multiple cause analysis and corrective actions for prevention of related occurrences) is not followed the program can fail. Using the program for discipline avoidance alone will not lead to a strong implementation. Labor must be open to providing detailed close call reports with sufficient information to enable analysis of the causes. On the other end of the process, labor also must work collaboratively with managers to implement corrective actions, lending their field expertise and demonstrating their support of the program to their peers. Management must also embrace the approach that learning about the contributing causes of close calls can help proactively prevent occurrences more than disciplining an individual. Management must devote time and resources to ensure corrective actions are implemented. Simply collecting close call data and doing nothing with it will produce few results.

However, with favorable implementation, three C^3RS sites experienced reduced derailments ranging between 20%–41%. This was the most common impact in the evaluation, suggesting that C^3RS shows promise in improving safety in transportation. It is also possible for C^3RS to help reduce discipline hearings and their associated cost and to reduce injuries.

For C^{3}RS to be sustained at a site, all the external and internal stakeholders must reach agreement on how to revise the IMOU, if needed. This requires agreement internally between labor and management and externally adding the FRA, NASA, and national labor. C^{3}RS can be sustainable in the railroad industry with both good implementation by individual carriers and continued support from FRA and national labor. So far, progress toward industry sustainability has been made with passenger railroads and not yet with freight railroads.

6.2 Validity of the Evaluation

This evaluation represents a large research project for FRA. The validity of this evaluation is a function of two elements: 1) the research design that underlays the evaluation and 2) the theory of change that underlays C^{3}RS as an innovation. The research design can be understood in terms of the logical structure of the design, the sampling scheme, the data used, and the tactics employed for data analysis and interpretation, as portrayed in Figure 13. Figure 9 articulates the theory of change, identifying the relationships in C^{3}RS structures and processes required if C^{3}RS is to have its desired consequences. The findings reported here are supported by multiple data sources, representing a convergence of evidence (GAO, 2009).

Research Design: The research design uses a quasi-experimental and mixed methods approach. When an experimental design with random assignment is not feasible, the most rigorous research design uses both quantitative and qualitative data collection and analysis methods from multiple data sources (GAO, 2009). The logical structure of research design used to address these questions is shown in Figure 13. As that model indicates, the C^{3}RS evaluation combined cross sectional and longitudinal observations:

- Intra-railroad comparisons
- Cross-railroad comparisons
- Analysis of change over time
- Many quantitative and qualitative data sources
- Balanced perspectives on the freight and passenger railroad industry

Due to the non-simultaneous initiation of the demonstration pilots at each railroad, the actual data collection spread over a 10-year period even though each demonstration's duration was only 5 years. As a result, it was possible to observe C^{3}RS through a range of business conditions and fluctuation in labor management relations. Thus, there is some confidence that the evaluation's findings are robust in the face of changing environmental conditions.

Multiple data sources included C^{3}RS reporting data, repeated interviews with all relevant stakeholders, the Evaluation Team's extensive field notes, validated surveys of safety culture, and corporate archival data. The mixed methods approach with a wide range of data types and coverage served two purposes. First, it often made it possible to increase confidence in findings by showing the same finding from multiple sources. Second, it broadened the range of analysis because often a particular data source was relevant to aspects of C^{3}RS behavior not covered by other sources.

The wide range of stakeholders afforded validity checks that paralleled the multiple data sources. In some cases, confidence in findings could be increased because that finding was confirmed by multiple stakeholder groups. In other cases, the wide range of stakeholders offered the

opportunity to observe C^{3}RS operating practices that could be understood only with respect to a unique stakeholder.

Theory of Change: The research design described above was used to test the logical chain that is expressed in C^{3}RS's theory of change (Figure 9). This provided an additional check on the validity of findings because no finding could be trusted if its required precursor activity did not exist. Another advantage of relying on the theory of change is that it identified hypotheses as to what aspects of C^{3}RS were important and how they related to each other, thus ensuring validity and utility with respect to understanding why C^{3}RS worked as it did, and how to improve it in the future.

At Site 2, the impacts were only significant if the Evaluation Team compared high to low usage times. This approach was needed because Site 2 did not have consistent levels of C^{3}RS activity over the course of their demonstration. Multiple measures had the same behavior of improving during high usage and reverting to baseline values during low usage. This approach allowed the Evaluation Team to determine how much impact C^{3}RS could have when fully implemented, as expressed in C^{3}RS's theory of change (Figure 9).

6.2.1 Limitations of the Evaluation

Despite the comprehensive nature of the evaluation, there are some limitations to take into account when drawing conclusions. The study involves four railroads at specific points in time. Whether other railroads might achieve similar results at other times would depend on whether the other railroads had similar characteristics to the four carriers in this study. Caution should be exercised when generalizing to Class 2 and Class 3 railroads since they were not represented in the four sites. Carrier stakeholders' reactions to C^{3}RS's Evaluation briefings provided some confidence that the results may be applicable to railroads similar to those included. In addition, FRA's enthusiasm in using the results of the evaluation in their recruiting efforts for new C^{3}RS participants is a positive indicator in the general applicability of the findings.

The Evaluation Team has the most confidence in the derailment reduction results observed at three of the four sites. The number of positive impacts on the same outcome is convincing. Site 3 did not reach the traditional accepted significance rate; however, the reduction in derailments due to C^{3}RS is plausible because Site 3 had evidence in their reports and corrective actions that they implemented several related corrective actions. Their results were heading in the right direction. Moreover, Site 3 had fewer reports related to derailments than Site 1 and 2, which could have translated into fewer derailment related corrective actions (there was missing corrective actions data from all sites) which could explain the not quite significant results. Using a medical analogy, Sites 1 and 2 had a "higher dose" of the medicine than at Site 3.

The study is limited in demonstrating impact beyond derailments, injuries, and disciplinary costs. C^{3}RS's impact on derailments was caused by the large number of derailment/switching-related reports generated across the sites. If more reports dealt with personal injuries, it is plausible that injury rates would decrease. However, because injury-related reports were few, there was minimal impact on injuries observed (only at Site 2).

The lack of reliable speed-related outcome measure for excess speed reports represents a limit of the evaluation. Excess Speed reports were the most common close call category. The evaluation was unable to determine if C^{3}RS was effective at helping to improve this safety critical issue.

However, the railroads did learn more about the frequency and causes of excess speeding, but it is unknown if the corrective actions they implemented were effective.

When questioned, some sites stated that their train miles volume was consistent from year to year, and thus decided not to provide the Evaluation Team with cars-moved or worker-hours to be used for normalization purposes in the data analysis. While they have insight into their own data, it is possible that they did not make perfect assumptions, and some variations may have occurred. Without the normalizing data, the Evaluation Team was not able to confirm their estimates.

Random sampling was not possible when the surveys were administered. Rather, as many of the effected workers as possible were asked to complete the survey and chose to cooperate. The high response rates might mitigate against the supposition that the respondents were not representative of their entire population. On the contrary, they did volunteer, and it is not unreasonable to assume some kind of self-selection bias.

Safety culture scale data was corroborated through the interviews with workers that were carried out at each site during baseline, midterm, and final periods. In preparation for these interviews, the Evaluation Team explicitly asked to include a variety of supporters and skeptics about C^3RS. And in fact, many interviewees were not shy about voicing negative opinions. Local managers chose the people interviewed from those who were at work on the day the data was collected, and it is possible that the managers were biased regarding who they selected to be interviewed. Also, because the evaluators were limited in their time at each site, the respondent pool was based on who was at work on a given day and available to participate in an interview.

An extensive cost-benefit analysis was not performed in this study. Interviewees talked about the PRT meetings being a significant cost, but complete and accurate information was not collected. Another difficulty in calculating the costs of different corrective actions was that the costs might have been spread across several departments. In terms of benefits, there is a potential benefit from reducing derailments, which can reduce the costs of repairing track and equipment. Reducing derailments could also provide benefits by increasing throughput and on-time schedule adherence. In spite of these challenges, the Evaluation Team did find labor cost savings due to the reduction in the number of discipline hearings.

The industry sustainability analysis was limited in four ways. 1) The sustainability data from demonstration sites were sparser than the implementation data. The Evaluation Team was only able to collect a few post-demonstration interviews, much less than the interviews conducted during implementation. As demonstration sites' personnel changed jobs, it was more difficult to find interviewees knowledgeable about C^3RS. The few available interviews and field notes revealed some of the factors influencing sustainability, but this is an area that could be researched further. 2) The evaluation scope did not include interviews with carriers that did not join C^3RS about their motivations and concerns. 3) The sustainability data analysis was more limited than the implementation analysis, because it relied solely on interview data, rather than the mixed methods approach used for implementation (e.g., adding corrective action and C^3RS reporting data). 4) There were also no shortlines in the scope of the evaluation. Given the restricted data available, there are limits on the confidence of the industry sustainability findings.

6.2.2 Alternative Explanations

6.2.2.1 Confounds

Given that this evaluation is a field study, it is inevitable that confounds and exogenous variables would emerge that could influence impacts observed. Sites 1 and 3 had other human factors-based safety programs operating for at least some of the time that C^{3}RS was operating. At Site 1, it was possible to perform analyses that provided some confidence that whatever the strength of the confounding program, C^{3}RS still had a discernable effect. That analysis was performed comparing results observed at Site 1 with a comparison site in the same service unit. Since the two sites operated in the same service unit, they reported to the same senior manager influencing some of the management personnel, policies, and practices. Further, the comparison site was similar in terms of the number of headcount and train volume. Thus, if Site 1 exhibited more improvement in safety than the comparison, it was plausible that the difference could be attributable to C^{3}RS. The demonstration did outperform the comparison site providing evidence that something unique was occurring at the treatment site.

Unlike Site 1, Site 3 did not have a comparison group for the derailment assessment; therefore, a potential confound with their other safety program limits the confidence of these results. There was no way to confirm that observations at the demonstration were unique. However, derailment reductions observed at the other two C^{3}RS demonstration sites strengthens confidence that C^{3}RS contributed to Site 3's improvement.

6.2.2.2 Effects of the Evaluation

The evaluation did not contain a rigorous assessment of its own role in the way that the various stakeholders engaged with C^{3}RS. However, two observations with respect to the sites and one with respect to the FRA are pertinent. Field note observations showed that the evaluation briefings given to the sites provided them with information on corrective actions and impacts that they themselves did not have. Thus, the evaluation compensated for a major weakness in how C^{3}RS was implemented which was the difficulty all sites had tracking the implementation progress and impact of corrective actions. Also, those briefings provided the sites with rich and detailed information about how C^{3}RS operated and how to improve those operations. The sites would not have this information with the Evaluation Team's presentations to them. In addition, interviews with HPD personnel revealed that C^{3}RS evaluation information was used in recruiting efforts for new railroads and for justifying funding in dealings with Congress thereby encouraging sustainability.

6.3 Moving Forward with C^{3}RS

The pilot demonstrations illustrated the potential for C^{3}RS to improve safety and safety culture. It also demonstrated the value that FRA safety programs can offer the industry.

6.3.1 Lessons Learned

To replicate C^{3}RS's implementation, impact, and sustainability successes, a railroad needs to engage in activities that were not deemed critical at the program start. A review of the findings reveals some lessons learned about C^{3}RS. Railroad management, labor, and FRA can use these lessons learned to achieve effective C^{3}RS processes at other sites.

6.3.1.1 Implementation

- **Detailed reports are more important than frequent reports.** When training eligible reporters on how to report to C^{3}RS, an emphasis on the detail of reports is critical. The PRT needs to be focused on detail too, and if there is not sufficient detail, the PRT needs to ask reporters what is getting in the way and address the barriers to detailed reporting. It is essential for determining contributing causes and improving safety.
- **Labor participating in corrective action implementation is essential to good implementation.** Labor should visibly show commitment to the program by leading and implementing corrective actions over which they have control. They need to cooperate with management as they are responsible for implementing corrective actions as well. Labor should use their expertise and relationships with their peers to ensure corrective actions are implemented successfully.
- **Management determining how to provide expertise for C^{3}RS data analysis is essential to getting the business case analysis on corrective actions.** The PRT analysis of close calls could benefit from the addition of cost-benefit analysis and risk management methods. Management should determine a way to connect analysis experts in their company with the C^{3}RS program. One possibility is to assign a person with those skills to the PRT. Another possibility is to involve people with those skills in the Support Team review of corrective actions. Either way, management needs to clearly communicate with the PRT about how advanced analysis functions will be accomplished with clear roles and responsibilities identified.
- **Managers should resolve disputes in a way that encourages future reporting.** When disputes occur over the discipline eligibility of a C^{3}RS reporter, it is tempting for managers to revert to the former culture of discipline. However, in order to protect the employees' trust in C^{3}RS, managers should resolve disputes in a manner that will increase, not decrease trust. All future reporting will be influenced by how disputes are resolved.
- **Personnel transitions can cause downtime and negatively affect bottom-line impact.** As PRT members from labor and management rotate, transitions should be carefully managed. This can be done by staggering rotations. For example, all union members should not rotate at the same time. Also, involving multiple managers helps mitigate the loss when one moves to a new position. If a PRT or Support Team member is planning to leave, replacements should be selected, trained, and involved ahead of time to ensure a smooth transition.
- **Sites can share common process improvements and corrective actions (that are not proprietary) with each other and the industry and thereby "lift all boats."** Across sites, similar close calls were reported, and similar corrective actions were implemented. FRA, NASA, and the participating sites should work together to find a way to collaborate with each other as well as inform the industry of similar contributing factors and solutions. With this type of knowledge sharing, "all boats could be lifted together." Continuing the User Group is one method to do this, but it does not communicate information to the rest of the industry that is not involved in C^{3}RS.

- **Sites have limited capacity to implement corrective actions; therefore, guidance on priorities can help target PRT analysis.** Sites are not realistically able to implement a corrective action to address every close call report or every recommendation from the PRT. The input of reports and recommendations will exceed the output of corrective actions. Sites need to have realistic expectations about how much can be implemented. Management needs to set priorities among the different options and communicate with the PRT about their decision making process. Data external to C^3RS may be utilized, such as the number of related incidents or injuries.
- **FRA provides significant value but may not be required at all PRT meetings.** During the demonstration, FRA participation on the PRT team was appreciated by the PRTs but was not considered necessary for every meeting. Even during the demonstration, it was sometimes difficult for FRA members to get budget to attend PRT meetings. As the program grows, budget concerns will grow. FRA may want to consider revised ways for FRA inspectors to be involved in C^3RS in the future.
- **Reasons railroads find it difficult to track corrective actions should be addressed to ensure more effective use of resources.** The finding showing that railroads find it difficult to track and publicize corrective actions implies that simply "trying hard" may be insufficient. The root and contributing causes for this weakness in the C^3RS process should be determined by each railroad. Each site needs to determine what tools and resources they have available.

6.3.1.2 Impact

- **Close call data are useful.** Railroads can discover safety problems that they did not know about through close call reporting. That knowledge can lead to corrective actions to help improve safety. Since railroads do not know exactly what issues are out there, the outreach needs to encourage employees to report about all sorts of close calls.
- **Carriers need to monitor corrective actions to determine if they are effective.** Carriers would benefit from setting targets for specific corrective actions, assigning a person responsibility for the action, and monitoring the applicable metrics. Both managers and labor from the PRT can contribute. This may involve going beyond the high level aggregated metrics to determine more sensitive measures that map to the details of their corrective actions (e.g., derailments caused by run-through switches). (Remember: official counts of run-though switching may increase as a result of the opening of communication from C^3RS and do necessarily mean that more run-though switches are actually occurring.)
- **Derailment close call reports seem to be useful in reducing derailments and encouragement could be helpful.** Carriers would benefit from encouraging their employees to submit close calls related to run-through switches and derailments. The success of the demonstration sites in reducing derailments showed that close call reports can contain sufficient information to find the causes of derailments.
- **Improvements in supervisor-employee relationships can be attained**. Allowing employees to receive discipline protection for submitting close call reports opens up communication with supervisors. When concern of discipline is removed, carriers may discover the employees are very willing to discuss safety concerns freely. Supervisors

should embrace this willingness to communicate and have productive, blame-free, conversations about safety with their employees.

6.3.1.3 Sustainability

- **FRA needs to determine ways to measure impact in the future to ensure continued justification and funding for the program.** FRA RRS relied on FRA RD&T's funding of the Evaluation Team to deliver impact results used to justify the program to new sites and obtain funding from FRA senior management and Congress. In the future, FRA RRS should determine how to continue to show that C^3RS is worthwhile.
- **Sustainability is strengthened by support from stakeholders external to a carrier.** Both FRA and national labor have roles to play to support carriers in their long-term sustainability. When disagreements over the process occur, external stakeholders should work closely with carriers and PRT teams to understand concerns and plan a path forward. The suitability of the C^3RS model should be continually reconsidered with respect to changes in technology, the needs of stakeholders, and conditions in the industry.
- **FRA should consider new approaches to reach out to Class 1 railroads.** At the moment, no Class 1 railroads, and only one shortline railroad, are involved in C^3RS. FRA should research the reasons why Class 1 railroads are not involved; consider possible changes to accommodate them; and implement new ways to reach out to them.
- **Sustainability is strengthened when success stories are shared.** Carriers within C^3RS should publicize the benefits of C^3RS to their workforce, local managers, and senior management. Newsletters and announcements about significant corrective actions demonstrate that C^3RS is finding and solving issues. Sharing stories about the benefit of C^3RS and objective metrics helps build support for C^3RS with senior management. If benefits are widely disseminated, then institutional knowledge of C^3RS is more likely to be maintained over time, even after personnel turnover. This may require improvements to the tracking of corrective action status, so it is clear which ones were actually implemented.

6.3.2 Implications

The substantial weight of evidence from this evaluation shows that railroads are capable of achieving improvements in safety and safety culture in the presence of C^3RS. It follows that the railroad industry and railroad workers would benefit from C^3RS implementation across the country. As of the writing of this report, FRA has already chosen to promote and support C^3RS implementations through the Human Performance Division by doing the following:

- **Awareness:** Raising awareness of C^3RS as a risk reduction strategy and its potential impacts through publicizing the results of this evaluation.
- **Support:** Continuing to support the national C^3RS User Group of FRA, labor officials, and railroad representatives.
- **Expertise:** Providing expertise for establishing good C^3RS implementation through education and training.

- **Funding:** Providing funding for NASA to be the Third-Party and for Human Performance Division staff to train and support C^3RS sites.
- **Communication**: Providing publications and briefings about the C^3RS evaluation results to the public and Congress.
- **Recruiting**: Discussing C^3RS with new carriers in an effort to increase industry participation in C^3RS.
- **Improvements**: Revising materials and software in an effort to better support C^3RS sites.

7. Conclusion

When the FRA C^{3}RS demonstration began, close call systems had already been implemented in other industries and countries. The U.S. railroad industry had unique challenges including a hundred-year-old history of contentious labor-management relationships and workman's compensation act that incentivized blame. Given these challenges, it was by no means certain that it could be implemented, and if implemented, that it would prove beneficial. This evaluation showed that C^{3}RS, if implemented well, can lead to bottom-line impacts. C^{3}RS demonstrated the ability to reveal unknown safety risks and their causes.

Three of the four demonstration sites achieved positive bottom-line impacts, and two sites sustained their participation in C^{3}RS long-term. Favorable implementation by carrier management and labor is essential but not sufficient. There must be common agreement with the external stakeholders on the future direction of the program as it evolves, perhaps being captured in revisions of the IMOU when tensions arise and a change is needed.

As C^{3}RS continues to evolve, there are opportunities to improve, including better communication; improved PRT data analysis and efficiency; and enhanced tracking of corrective actions. Further research could be conducted to determine what modifications to C^{3}RS might help to increase interest on the part of freight railroads.

References

Aviation Safety Reporting System (ASRS).

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Lawrence Erlbaum Associates.

Colquitt, J. A. (2001). On the Dimensionality of Organizational Justice: A Construct Validation of a Measure. *Journal of Applied Psychology, 86*(3), 386–400.

Coplen, M. K. (1999). "Compliance with Railroad Operating Roles and Corporate Culture Influences: Results of a Focus Group and Structured Interviews." Technical Report, DOT/FRA/ORD-99/09. Washington, DC: Federal Railroad Administration.

Creswell, J. W. (2014). A Concise Introduction to Mixed Methods Research. SAGE Publications, Inc.

Dastmalchian, A., Blyton, P., and Adamson, R. (1989). Industrial Relations Climate: Testing a Construct. *Journal of Occupational Psychology, 62*(1), 21–32.

Davidson, J. E. (2005). Evaluation Methodology Basics. Thousand Oaks, CA: Sage Publications

Dedobbeleer, N., and Beland, F. (1991). A safety Climate Measure for Construction Sites. *Journal of Safety Research, 22*(2), 97–103.

Eisenberger, R., Huntington, R., Hutchison, S., and Sowa, D. (1986). Perceived Organizational Support. *Journal of Applied Psychology, 71*(3), 500–507.

Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I. L., and Rhoades, L. (2002). Perceived Supervisor Support: Contributions to Perceived Organizational Support and Employee Retention. *Journal of Applied Psychology, 87*(3), 565–573.

Federal Railroad Administration. (2012). "Senior Cross-functional Support—Essential for Implementing Corrective Actions at C^{3}RS Sites." Research Results, RR12-09. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2013). "Another C^{3}RS Site Improves Safety at Midterm." Research Results, RR 13-49. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2014). "Update from C^{3}RS Lessons Learned Team: Four Demonstration Pilots." Research Results, RR 14-17. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2014). "Update from C^{3}RS Lessons Learned Team: Safety Culture and Trend Analysis." Research Results, RR 14-18. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2015). "Continued Improvements at One C^{3}RS Site." Research Results, RR 15-17. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2016). "C^{3}RS: Midterm Accomplishments at Another Site and Success Factors Across Sites." Research Results, RR 15-44. Washington, DC: Federal Railroad Administration.

Federal Railroad Administration. (2016). "C^{3}RS Is Implementing Corrective Actions and Expanding Within the Railroad Industry." Research Results, RR-08. Washington, DC: Federal Railroad Administration.

Greene, J. C. (2007). Mixed Methods in Social Inquiry. Jossey-Bass.

Hofmann, D. A. (2004). Safety Climate: Management Attitudes Measure [personal communication].

Hofmann, D. A., and Stetzer, A. (1998). The Role of Safety Climate and Communication in Accident Interpretation: Implications for Learning from Negative Events. *Academy of Management Journal, 41*(6), 644–657.

Hofmann, D. A., and Stetzer, A. (1996). A Cross-Level Investigation of Factors Influencing Unsafe Behaviors and Accidents. *Personnel Psychology, 49*(2), 307–339.

Hofmann, D. A., and Morgeson, F. P. (1999). Safety-Related Behavior as a Social Exchange: The Role of Perceived Organizational Support and Leader-Member Exchange. *Journal of Applied Psychology, 84*(2), 286–296.

Ishikawa, K (1982). Guide to Quality Control. Asian Productivity Organization. Tokyo, Japan

Johns, M. W. (1992). Reliability and Factor Analysis of the Epworth Sleepiness Scale. *Sleep: Journal of Sleep and Sleep Disorders Research, 15*(4), 376–381.

Lee, E. T., and Wang, J. (2003). *Statistical Methods for Survival Data Analysis*. Hoboken, N.J.: John Wiley & Sons.

Lewin, K. (1947). Group Decision and Social Change. In T. N. Newcomb and E. L. Hartley (eds.) *Readings in Social Psychology*. Troy, MO: Hold, Rinehart, and Winston.

Magley, V. J., Davies-Schrils, K., and Walsh, B. (2007). Workplace Climate for Incivility: Development of a New Measure. Unpublished manuscript, University of Connecticut.

Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. 1952. The Time Intervals Between Industrial Accidents. *Biometrika*, *39*(1–2), 168–180.

Morell, J. A., Hanssen, C., Thompson, D., Wallace, R., and Wygant, B. (2006). Confidential Close Call Reporting in the Railroad Industry: A Literature Review to Inform Evaluation. Washington, DC: Federal Railroad Administration.

Mueller, L., Da Silva, N., Townsend, J. C., and Tetrick, L. (1999). An Empirical Evaluation of Competing Safety Climate Measurement Models. Paper presented at the Annual Meeting of the Society for Industrial and Organizational Psychology. Atlanta, GA.

National Research Council. (1994). Compensating Injured Railroad Workers Under the Federal Employer's Liability Act: Special Report 241. Washington, DC: The National Academies Press.

Naumann, S. E., and Bennett, N. (2000). A Case for Procedural Justice Climate: Development and Test of a Multilevel Model. *Academy of Management Journal, 43*(5), 881–889.

Nelson, W. B. (1982). *Applied Life Data Analysis.* Hoboken, NJ: John Wiley & Sons.

Nelson, W. B. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. Philadelphia: Society for Industrial and Applied Mathematics.

Niehoff, B. P., and Moorman, R. H. (1993). Justice as a Mediator of the Relationship Between Methods of Monitoring and Organizational Citizenship Behavior. *Academy of Management Journal,* 36, 527–556.

NJ Transit website. About Us. Available at: http://www.njtransit.com/tm/tm_servlet.srv?hdnPageAction=CorpInfoTo.

Organ, D., and Konovsky, M. (1989). Cognitive Versus Affective Determinants of Organizational Citizenship Behavior. *Journal of Applied Psychology, 74*(1), 157–164.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (1987). *How to Use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage.

Petty, N. W. (2016). Oh Ordinal data, what do we do with you?.

Ranney, J., and Nelson, C. (2004). Impacts of Participatory Safety Rules Revision in U.S. Railroad Industry: An Exploratory Assessment. *Journal of the Transportation Research Board,* No. 1899, Transportation Research Board of the National Academies, Washington, DC, pp. 156–163.

Ranney, J., Morell, J., Davey, M., Morell, and J., Raslear, T. (2015). "Confidential Close Call Reporting System (C^3RS) Lessons Learned Team Baseline Phase Report." Technical Report, DOT/FRA/ORD-15/08. Washington, DC: Federal Railroad Administration.

Ranney, J., and Raslear, T. (2012). "Derailments Decrease at a C^3RS Site at Midterm." Research Results, RR12-04. Washington, DC: Federal Railroad Administration. Available at:.

Raslear, T., Ranney, J., and Multer, J. (2008). "Confidential Close Call Reporting System: Preliminary Evaluation Findings." Research Results, RR08-33. Washington, DC: Federal Railroad Administration.

Reason, J. (1998). Safety Culture: Some Theoretical and Practical Issues. *Work and Stress*. *12*(3), 293-310.

Robinson, S. L. (1996). Trust and Breach of the Psychological Contract. *Administrative Science Quarterly, 41*(4), 574–599.

Rogers, E. M. (1995). *Diffusion of Innovations*. New York: Free Press.

Rosenthal, R., and Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis* (2nd ed.). New York: McGraw-Hill.

Rossi, P. H., Freeman, H. E., and Lipsey, M. W. (1999). *Evaluation: A Systematic Approach* (6th ed.). Thousand Oaks, CA: Sage.

Saks, J., Multer, J., Blythe, K. (2004). Proceedings of the Human Factors Workshop: Improving Railroad Safety Through Understanding Close Calls. Department of Transportation: Federal Railroad Administration: Office of Research and Development. (DOT-VNTSC-FRA-04-01)..

Schein, E. H. (1992). *Organizational Culture and Leadership* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Scriven, M. (1991). *The Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston: Houghton Mifflin.

Simard, M., and Marchand, A. (1997). Workgroups' Propensity to Comply with Safety Rules: The Influence of Macro-Micro Organizational Factors. *Ergonomics, 40*(2), 172–188.

Sullivan, C. (2001). Who Cares About CAIR? Australia and New Zealand Society of Air Safety Investigators. Cairns, Queensland.

U.S. Government Accountability Office. (2009, November). Program Evaluation: A Variety of Rigorous Methods Can Help Identify Effective Interventions. Washington, DC: U.S. Government Accountability Office.

Title V – Confidential Information Protection and Statistical Efficiency. (2002, December 17). 116 Stat. 2962. Public Law 107–347.

Transportation Review Board. (1994). Compensating Injured Railroad Workers Under the Federal Employers' Liability Act. Issue 241.

Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*. New Series, *185*(4157), 1124–1131.

Wayne, S. J., Shore, L. M., and Liden, R. C. (1997). Perceived Organizational Support and Leader-Member Exchange: A Social Exchange Perspective. *Academy of Management Journal, 40*(1), 82–111.

Weisberg, H. I. (2014). *Willful Ignorance: The Mismeasure of Uncertainty*. Wiley, NY, pp. 294–298.Winer, B. J., Brown, D. R., and Michels, K. M. (1991). *Statistical Principles In Experimental Design* (3rd ed.). New York: McGraw Hill.

Yin, R. K. (2014). Case Study Research: Design and Methods 5th ed. Thousand Oaks, CA: Sage.

Zohar, D. (1980) Safety Climate in Industrial Organizations: Theoretical and Applied Implications. *Journal of Applied Psychology, 65*(1), 96–102.

Zuschlag, M., Ranney, J., Coplen, M., and Hamar, M. (2012). "Transformation of Safety Culture on the San Antonio Service Unit of the Union Pacific Railroad." Technical Report, DOT/FRA/ORD-12/16.

45 U.S.C. § 51 et seq. (1908).

Abbreviations and Acronyms

Abbreviations	Acronyms
AEA	American Evaluation Association
ATDA	American Train Dispatchers Association
ASRS	Aviation Safety Reporting System
BLET	Brotherhood of Locomotive Engineers and Trainmen
BTS	Bureau of Transportation Statistics
CP	Canadian Pacific Railway
CFR	Code of Federal Regulations
C^3RS	Confidential Close Call Reporting System
CIPSEA	Confidential Information Protection and Statistical Efficiency Act
CI	continuous improvement
FELA	Federal Employers Liability Act
FRA	Federal Railroad Administration
FOUO	For Official Use Only
HF	Human Factors Division
IMOU	Implementation Memoranda of Understanding
SMART	International Association of Sheet Metal, Air, Rail and Transportation Workers
MOU	Memorandum of Understanding
MCIA	multiple cause incident analysis
MANACOVA	multivariate analysis of covariance
MANOVA	multivariate analysis of variance
NDA	nondisclosure agreement
NASA	National Aeronautics and Space Administration

Abbreviations	Acronyms
NTSB	National Transportation Safety Board
NJT	New Jersey Transit
RRS	Office of Railroad Safety
RD&T	Office of Research, Development and Technology
PRT	Peer Review Team
SMART TD	SMART Transportation Division (formerly UTU)
TCU	Transportation Communications Union
TRB	Transportation Review Board
UP	Union Pacific Railroad
UTU	United Transportation Union (later called SMART TD)
Volpe Center	Volpe National Transportation Systems Center

Glossary

Close Call – an opportunity to improve safety practices in a situation or incident that has a potential for more serious consequences (Saks, 2004).

Close Call category – The category of the close call event, as identified by the C^{3}RS Third-Party (e.g., Excess Speed, Yard – switch, Yard – derail, etc.) (Categories were assigned by BTS or NASA).

Close Call Event – a specific situation or event that has a potential for more serious consequences that was observed by a railroad employee.

C^{3}RS – Confidential Close Call Reporting System – an FRA sponsored voluntary confidential program allowing railroad carriers and their employees to report close calls. The program provides a safe environment for employees to report unsafe events and conditions. Employees receive protection from discipline and FRA enforcement. Railroads also receive protection from FRA enforcement for events reported within C^{3}RS.

C^{3}RS-Analyzed-Case – contains the C^{3}RS-report-record sent from the C^{3}RS Third-Party to the PRT and the results from the PRT's analysis. Each C^{3}RS-analyzed-case contains a close call category assignment that was provided within the report-record.

C^{3}RS Demonstration Pilot Sites – the sites at the first four railroads to join the C^{3}RS program. They joined during what FRA considered the demonstration phase. During this time FRA was evaluating whether or not to expand C^{3}RS and open it up to other railroads. Each demonstration pilot site committed to participate in C^{3}RS for 5 years in a portion of their railroad.

C^{3}RS Evaluation Team – the group of evaluation experts assembled by the Volpe National Transportation Systems Center to ascertain how well C^{3}RS was working, what it accomplished, and what would be needed to sustain it over the long run.

C^{3}RS Implementation Team – the group of personnel from the Volpe National Transportation Systems Center and later from FRA RRS HPD who planned the implementation of the C^{3}RS at the demonstration sites, trained the PRT, and provided some support to the sites.

C^{3}RS Report – is created when an individual railroad employee observes a close call event and either makes a phone call or electronic submission to the Third-Party, and it is accepted under the criteria laid out in that railroad's IMOU.

C^{3}RS-Report-Record – contains de-identified information from the C^{3}RS Third-Party about a single close call event. It can originate from a single employee's C^{3}RS report or several employees' reports about the same close call event. The Third-Party consolidates the information from the original report(s), conducts follow-up interviews with the reporter(s) as available, removes identifying information, assigns it to a close call category, and creates a written C^{3}RS-report-record. Then, the Third-Party sends report-records to the PRT for analysis.

C^{3}RS Site – the generic name for any railroad that is implementing the C^{3}RS program.

C^{3}RS Third-Party – the agency that collects C^{3}RS reports, completes interviews with the reporting employee, combines reports on the same close call event from multiple employees, redacts identifying information, and forwards the information to the relevant railroad.

Corrective Actions – an action that a railroad can take to mitigate a safety risk. The PRT creates recommendations for corrective actions based on their C^{3}RS case analysis. Corrective actions can be specific to a particular location, or applicable to a wider geographical area and/or across organizational boundaries.

Hazard – any source of potential damage, harm, or adverse health effects on something or someone under certain conditions at work.[40]

Human Factor Caused Accident – an accident that is directly attributable to the operator, worker, or personnel involved in an accident.[41]

Known Close Call Event – an event that is below the FRA reporting threshold for operating rules and does not involve an injury, but would require managerial notification if discipline protection was sought through C^{3}RS. To facilitate analysis of such events, employees provide notification of the event to management without undue delay in addition to a C^{3}RS report. The scope of known events eligible for C^{3}RS protection from discipline is determined by the railroad's IMOU in Section 6.4.

Lessons Learned Evaluation – the evaluation carried out by the C^{3}RS Evaluation Team.

Multiple Cause Incident Analysis (MCIA) – the method for analyzing data used by the PRT. MCIA consists of a series of questions to help identify multiple contributing factors that triggered a close call event. An electronic MCIA tool was used to track and record this work.

Peer Review Team (PRT) – a team at made up of management, labor, and FRA at a single C^{3}RS site that receives data from the C^{3}RS Third-Party, analyzes it, and provides recommendations to local and corporate management.

Support Team – a group of managers at a C^{3}RS site that is responsible for reviewing corrective action recommendations from PRT and implementing them as needed.

Reportable Incidents – an incident, such as a derailment or collision, whose cost exceeds the amount established by FRA, and is thus reportable to the agency.

Rubric – A rubric is a four point scale was used to rate each factor—very good, good, fair, and poor. Following the tradition of using scoring rubrics in the field of education to assess students, precise definitions for each of these ratings were developed, and continually referred to by the Evaluation Team as each factor in the Ishikawa model was assessed with respect to each demonstration site.

Safety Culture – refers to the ways that safety issues are addressed in a workplace. One expression of safety culture is "the attitudes, beliefs, perceptions, and values that employees share in relation to safety." Or in other words, simply "the way we do safety around here."[42]

---

[40] Canadian Center for Occupational Health and Safety. OSH Answers Fact Sheets.

[41] Safeopedia. Human Factors Causing Accidents.

[42] Wikipedia. Safety culture.

Appendix A. Evaluation Standards Attestation Form

In 2013, the Federal Railroad Administration Office of Railroad Policy and Development created an Evaluation Implementation Plan as a foundation for guiding systematic, improvement-oriented evaluations and institutionalizing program evaluation throughout the Office of R&D. While the C^{3}RS Evaluation began in 2006, before the Evaluation Implementation Plan was created, it does comply with the standards as described in the C^{3}RS Evaluation Standards Attestation Form in this appendix. The summaries of ANSI-approved standards were drawn from the Joint Committee on Standards for Educational Evaluation and have been adopted for use by FRA's RD&T.

Form Instructions:

> "Evaluators of Research and Development (R&D) programs should complete a copy of this form and append it to their final report, as an attestation of the extent to which the evaluation adhered to applicable, specific standards of Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability. This is reprinted from FRA's *Evaluation Implementation Plan* of 21 November 2013 and a Draft of the Attestation Form of 22 August 2013[43].
>
> The following summaries of ANSI-approved standards—drawn from Joint Committee on Standards for Educational Evaluation (2011). *The Program Evaluation Standards.* Los Angeles, CA: Sage.—are reprinted with the Committee's authorization and have been adopted for use by FRA's Office of R&D.[44]"

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
U1 Evaluator Credibility	*Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.*	The evaluation was conducted by experienced, independent evaluators. Evaluators came from Volpe's Surface Transportation Human Factors Division (Joyce Ranney and Cassandra Cantu). Outside experts were contracted to assist, Dr. Jonathan Morell (Syntek Technologies Inc.) and Melinda Davey (Jacobs). These evaluators together have a strong familiarity with FRA RD&T HF, C^{3}RS, and extensive experience in program evaluation.	**X**			

[43] The designations U, F, P, A, and E, respectively refer to categories of standards labeled Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability, detailed in EVALUATION IMPLEMENTATION PLAN: Office of Research & Development" FRA/ OSD (Nov 2014). DOT/FRA/ORD-13/47. Appendix A.8.

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
U2 Attention to Stakeholders	*Evaluations should devote attention to the full range of individuals and groups invested in the program or affected by the evaluation.*	The evaluation included interviews with all core stakeholders, including the FRA RD&T HF and FRA RRS RRP and HPD, the four demonstration pilot rail carriers (UP, CP, NJT, and Amtrak), C^3RS Implementation Team, BTS, and NASA. In addition, all these groups were frequently briefed on the evaluation findings.	**X**			
U3 Negotiated Purposes	*Evaluation purposes should be identified and revisited based on the needs of stakeholders.*	The original design of the evaluation was crafted to meet the needs of stakeholders who were risking real and political capital on a program with uncertain success. As the evaluation proceeded, analyses were often done to answer questions posed by these stakeholders.	**X**			
U4 Explicit Values	*Evaluations should clarify and specify the individual and cultural values underpinning the evaluation purposes, processes, and judgments.*	Evaluation purposes, processes, and judgments were framed by the underlying value of efficient and effective use of government resources in service of the public good.	**X**			
U5 Relevant Information	*Evaluation information should serve the identified and emergent needs of intended users.*	The evaluation findings and recommendations were designed with the practical intent of sharing lessons learned with the railroads and FRA to help inform decisions about the future of C^3RS, improve its implementation, and increase long-term sustainability. Findings were sought and enthusiastically embraced by all the stakeholders. Stakeholder input was used to guide plans for new analysis at the project proceeded. This was accomplished through multiple stakeholder engagement methods as described in Section 3.4.2, including but not limited to demonstration pilot site feedback sessions, C^3RS User Group Meetings which included new sites, and outsider reviews by the C^3RS Implementation Team and FRA.	**X**			
U6 Meaningful Processes and Products	*Evaluation activities, descriptions, findings, and judgments should encourage use.*	The evaluation deliverables undergo many reviews to make sure they are understandable and useful. C^3RS Implementation Team Members, Jordan Multer and Jane Saks, and FRA RD&T HF and RRS HPD reviewed deliverables. Also presentations were tailored for specific audiences and their needs, for example presentations to railroad workers have different content than presentations to executives. One of the functions of the reviews was to assure careful attention by the stakeholders doing the reviewing. Use was also encouraged by the demonstration sites providing feedback when they though a result was correct or incorrect.	**X**			

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
U7 Timely and Appropriate Communicating and Reporting	*Evaluations should attend in a timely and ongoing way to the reporting and dissemination needs of stakeholders.*	Frequent feedback to stakeholders was a key part of the evaluation. Each railroad received a tailored presentation after each of their phases of the evaluation: baseline, midterm, and final. These presentations included discussions with the labor, management, and FRA stakeholders to assure that understood the findings, verify the accuracy of the evaluation conclusions, and make plans for actions for the stakeholders to take in the future to improve their C^3RS programs. Further, many reports of findings have been published as FRA Research Results on their website. Other stakeholders receive at least yearly communications at User Groups, meetings between FRA and the Evaluation Team, and conferences such as the Transportation Review Board.	**X**			
U8 Concern for Consequences and Influence	*Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.*	Summary evaluation findings were concisely described and published in publically available FRA Research Results. The intent was to have credible, published sources of information on C^3RS lessons Learned. FRA RD&T reviewed all publications and Volpe project managers were involved.	**X**			
F1 Project Management	*Evaluations should use effective project management strategies.*	Evaluators developed and followed an evaluation plan. Contractors used an established system for internal project management and reported status against deliverables and milestones through Monthly Progress Reports, while Volpe provided information to FRA. The team (Volpe, Jacobs, and Syntek Technologies Inc.) also met weekly to review current activities and deliverables. The status of deliverables was tracked in a master spreadsheet.	**X**			
F2 Practical Procedures	*Evaluation procedures should be practical and responsive to the way the program operates.*	The evaluation developed feasible recommendations based on input from experts and project stakeholders familiar with the project and railroad operating environments. Data collection procedures were carefully negotiated at each demonstration pilot site to assure that logistics and data burdens were within limits that were acceptable to each site.	**X**			
F3 Contextual Viability	*Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.*	From the beginning, and over the course of the evaluation, the evaluators were keenly aware of the different points of view held by the government, the industry, and labor. There was a continual effort to address the unique information needs of each.	**X**			
F4 Resource Use	*Evaluations should use resources effectively and efficiently.*	The evaluation complied with Volpe Center's project management and status reporting practices. Contractor rates were set using a GSA schedule. Work has been done on budget and on schedule.	**X**			

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
P1 Responsive and Inclusive Orientation	*Evaluations should be responsive to stakeholders and their communities.*	Railroad's Peer Review Teams were shown their site findings first, so feedback could be collected and adjustments made before findings were shared outside that railroad.	**X**			
P2 Human Rights and Respect	*Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.*	Individual stakeholder information was kept confidential. Interviewee names were not recorded with the notes to protect their personal privacy. Personally Protected Information (PPI) was not included in the railroad safety data requests. The Volpe Center and the contractors created data protection plans and trained staff in their use. At all appropriate times, provisions of the Confidential Information Protection and Statistical Efficiency Act were followed, as were procedures for keeping information "company confidential." All data were stored in secure, encrypted form.	**X**			
P4 Clarity and Fairness	*Evaluations should be understandable and fair in addressing stakeholder needs and purposes.*	In scoping the evaluation, the evaluation questions, findings, and recommendations were designed to balance the interests of project stakeholders. Presentations were tailored for specific audiences. Both strengths and weaknesses of the program were included in the reports.	**X**			
P5 Transparency and Disclosure	*Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.*	Evaluation deliverables include a baseline, midpoint, and final technical report, containing a detailed description of the methods employed and overall findings. These reports will be publicly available and shared with project stakeholders. The plan also contains the publication of over 10 FRA Research Results over the duration of the evaluation to share findings in a more concise and frequent manner.	**X**			
P6 Conflicts of interests	*Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.*	The evaluation clearly documents funding sources. It also includes reviewers from many stakeholders to assure the results were not biased (Section 3.4.2).	**X**			
P7 Fiscal Responsibility	*Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.*	The evaluation complied with Volpe Center's IAA project management and status reporting practices.	**X**			
A1 Justified Conclusions and Decisions	*Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequence.*	A range of disciplines and organizational perspectives were taken into account in the design of the evaluation and conclusions were reviewed by a diverse set of stakeholders. As a result the Evaluation Team is assured that analyses were presented in a way that all could understand, and that the findings were relevant for them.	**X**			

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
A2 Valid Information	*Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.*	The draft reports are all reviewed and validated by the stakeholders familiar with the project as outlined in Section 3.4.2. Technical reviews assured that the information conveyed was credible.	**X**			
A3 Reliable Information	*Evaluation procedures should yield sufficiently dependable and consistent information for the intended use.*	The draft reports are all reviewed and validated by the stakeholders. Technical reviews assured that the information conveyed was credible.	**X**			
A4 Explicit Program and Context Descriptions	*Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.*	A description of the project, its history, and goals are included in all reports and documents that are disseminated to the public.	**X**			
A5 Information Management	*Evaluations should employ systematic information collection, review, verification, and storage methods.*	For interviews, consistent interview protocols were used. Interviews were analyzed using a consistent set of interview codes. For safety and survey data, similar statistical test were employed as were applicable by the types of data collected. Survey protocols were consistent. As the data collection progressed to new railroads and phases, the Evaluation Team used learnings from past data collection and analysis efforts to inform the later collection and analysis. The review process for deliverables was also consistent. Also lessons, methods, and documentation from other Volpe Center projects were shared with the C^3RS team to ensure consistency between Volpe Center evaluation projects. The railroad culture survey was based on published and validated scales and has been used in past Volpe Center projects.	**X**			
A6 Sound Designs and Analyses	*Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes*	The Evaluation Team included Dr. Jonathan Morell, an expert and active member of the evaluation field, who participated in the planning and execution of the evaluation plan. Melinda Davey, MSE in industrial engineering, conducted the initial data analysis based on team discussions and past experience. Volpe Center member, Michael Zuschlag, a research design and methodologist expert, and Wayne Nelson, an outside statistical expert, reviewed the technical methodology and findings.	**X**			
A7 Explicit Evaluation Reasoning	*Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.*	The evaluation is divided into clear sections – data, findings, recommendations – and there are clear links between the themes in each section. Extensive review for clarity and ease of reading was carried out.	**X**			

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
A8 Communication and Reporting	*Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.*	The evaluation included interviews and review of documents with representatives from the stakeholders (Section 3.4.1).	**X**			
E1 Evaluation Documentation	*Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.*	The evaluation includes documentation of the purpose and methodologies employed. All reports fully explain the history of the program, why evaluation was done, how stakeholders were involved, why particular analyses were done, and why specific conclusions were drawn.	**X**			
E2 Internal Metaevaluation	*Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.*	Volpe member, Michael Zuschlag, performed an internal (to Volpe) review of the technical methodology and findings related to corporate safety data. The Railroad Safety Culture Survey validated for use in other Volpe evaluations based on literature was used for C^3RS. Methods for data analysis and reporting formats from other Volpe studies were considered during the evaluation.		**X**		
E3 External Metaevaluation	*Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluations using these and other applicable standards.*	Wayne Nelson, an outside statistical expert, performed an external (to Volpe) review of the technical methodology and findings related to corporate safety data. Involved stakeholders encouraged the evaluators' policy of disseminating results with the involved demonstration sites to help with their process improvement and ensure the Evaluation Team's understanding. Once lessons learned were clarified they were shared publicly with the demonstration sites' approval.		**X**		

Appendix B. List of Phased Site Interview Questions

OMB CONTROL NUMBER: 2130-0574

RAILROAD EMPLOYEES VIEWS OF C^{3}RS

Paperwork Reduction Act Burden Statement

A Federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with a collection of information subject to the requirements of the Paperwork Reduction Act unless that collection of information displays a currently valid OMB Control Number. The OMB Control Number for this information collection is **2130-0574**. Public reporting for this collection of information is estimated to be approximately 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, completing and reviewing the collection of information.

All responses to this collection of information are voluntary. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: Information Collection Clearance Officer, Federal Railroad Administration, 1200 New Jersey Avenue, SE, Washington, D.C. 20590.

Introduction

One of the objectives of the C^{3}RS Evaluation Team is to determine what is required to improve the way C^{3}RS is implemented. We need this information to make recommendations for future implementations of the program. This interview is part of that effort. It will take about half an hour. I am only interested in the how C^{3}RS is going, not the substance of reports. To protect individual's privacy, we are not recording any names. All we need is a general description of respondents, e.g., "member of PRT; labor or management." Your participation in this interview is voluntary. If you want to skip any questions, please let us know. Thank you for meeting with us.

C^{3}RS (if labor)

L-1: Have you heard of C^{3}RS? (if No skip to S-1)

L-2: Do you think you understand the C^{3}RS well enough to know a reportable close call if you saw one?
Probe: What kinds of events have you been told can be reported?

L-3: Have you submitted a C^{3}RS report?

L-3a: (If they submitted a report) What did you think of your experience with the reporting system and BTS/NASA?

L-4: Do you know if C^{3}RS has resulted in any changes at your railroad?

L-5: Please tell us what changes you have seen.

L-6: How did you find about that these changes were made?

C^{3}RS (if manager)

M-1: Have you heard of C^3RS? (If No skip to S-1)

M-2: Do you think you understand the C^3RS well enough to give advice to your employees about what to report?

M-3: Do you know if C^3RS has resulted in any changes at your railroad?

Probes

- What are the changes?
- How did you find out about them? (formal vs. informal communication)
- Impact on
  - Safety culture: How management and labor interact.
  - Safety awareness
  - Safety (incidents, injuries, decertification)
  - Discipline
  - Cost

M-4: Have you personally been involved implementing any C^3RS corrective actions?

C^3RS (all)

A-1: From what you have seen of C3RS, what changes would you suggest to make it work better or be more effective in improving safety?

A-2: To what extent do you think management is supportive of C^3RS?

A-3: To what extent do you think labor officials are supportive of C^3RS?

A-4: To what extent are your friends and colleague supportive of C^3RS?

A-5: If you had to bet $5.00, would you bet that C^3RS will be up and running at UP in five years? Why?

Safety in general (leave out if running out of time)

S-1: Over the past year or so have any safety initiatives taken place other than C^3RS?

S-1a:- Do they overlap or interact with C^3RS?

S-1b:- Do you think that C^3RS can improve safety in ways that other safety programs can't?

S-2: How would you describe the average worker's attitude about safety at your railroad?

S-3: How would you describe management's attitude about safety at your railroad?

S-4: How would you describe labor management relations regarding safety at your railroad?

S-5: How would you describe relations between labor and management regarding issues other than safety at your railroad?

S-6: Have relations between labor and management changed over the past year?

Appendix C. List of Implementation Interview Questions

OMB CONTROL NUMBER: 2130-0574

Paperwork Reduction Act Burden Statement

A Federal agency may not conduct or sponsor, and a person is not required to respond to, nor shall a person be subject to a penalty for failure to comply with a collection of information subject to the requirements of the Paperwork Reduction Act unless that collection of information displays a currently valid OMB Control Number. The OMB Control Number for this information collection is **2130-0574**. Public reporting for this collection of information is estimated to be approximately 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, completing and reviewing the collection of information.

All responses to this collection of information are voluntary. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: Information Collection Clearance Officer, Federal Railroad Administration, 1200 New Jersey Avenue, SE, Washington, D.C. 20590.

C^{3}RS Implementation Interview Protocol (OMB No. 2130-0574)

Introduction

One of the objectives of the C^{3}RS Evaluation Team is to determine what is required to improve the way C^{3}RS is implemented. We need this information to make recommendations for future implementations of the program. This interview is part of that effort. It will take about half an hour. I am only interested in the how C^{3}RS is going, not the substance of reports. To protect individual's privacy, we are not recording any names. All we need is a general description of respondents, e.g., "railroad name, member of PRT; labor or management." Your participation in this interview is voluntary. If you want to skip any questions, please let us know. Thank you for meeting with us.

1-Thinking back over the past three months, what are the two or three most important positive or negative events that affected C^{3}RS?

Probes after description for each issue:

1-Why was this event so important?

2-Why do you think this event showed up when it did?

2-How satisfied are you with how C^{3}RS is currently working, and why do you feel that way?

2a. Probes:

- Peer Review Team
- Support Team Activities

2b. How could any of the groups involved in C^{3}RS change to improve C^{3}RS?

Listen, probe as necessary:

- Local management
- Local labor
- Railroad senior management
- BTS
- NASA
- FRA
- National labor

2c. Are there corrective actions that have been implemented that you think could have a big impact on safety? In addition to the ones you have mentioned, what are the kinds of corrective actions that are being implemented?

2d. Has C^3RS had any impact?

3-Over the past few months, are there any important events that took place outside normal C^3RS activities that affected the implementation or running of C^3RS?

4-Are there any issues effecting C^3RS's ability to maintain itself in the long run?

Appendix D. List of Qualitative Data Codes

This list contains codes for the qualitative data. They are organized by the five areas of the logic model. Not all of these codes had significant frequency.

Code Family: Implementation

- 6.4 as implementation motivator
- Communication with all stakeholders
- Credibility of key members
- Groups opposing implementation
- Implementation start-up
- Innovation champion
- Involving other crafts
- Key start-up meetings
- Local representation in early implementation
- Outreach to workers
- Past experience with change/collaboration
- PRT initial operations
- Signing IMOU

Code Family: Operations

- BTS activities
- C^3RS proves itself
- C^3RS reporting
- Change in PRT – company interface
- Confidentiality maintained
- Cross functional involvement
- Data quality
- Expanding C^3RS
- FRA participation on PRT
- Implementation of corrective actions
- Irrelevant agendas
- Learning curve

- NASA activities
- Poor participation
- PRT analysis
- PRT meetings
- PRT process experts
- PRT Support Team activities
- PRT tools
- PRTs sharing info
- Steering Committee/dispute resolution activities
- Tracking corrective actions

Code Family: Impact

- Culture change
- Employee engagement
- Impact
- Impact on FRA
- Impact on productivity
- Impact on safety
- Information-sharing among railroads
- No impact
- Reporting impact

Code Family: Environment and Internal Climate/Culture

- Age differences
- Company attitude toward safety
- Competition
- Conflict among members
- Conflicted position of low-level managers
- Contradictory corporate policies
- Differences among railroads
- Discipline vs. cooperative approach to safety
- FRA regulations - hours of service
- Groups vs. individuals
- Image of C^3RS among workforce
- Labor attitude toward safety
- New discipline policy
- New railroads
- Other accidents
- Other close call reporting programs outside railroads
- Passenger vs. freight
- Personal responsibility
- Safety problems
- Safety programs — other
- Safety vs. money
- Sharing track with other railroads
- Support from corporate
- Support from FRA
- Support from managers
- Support from NASA
- Support from NJT state
- Support from Steering Committee
- Support from unions
- Weather interference

Code Family: Sustainability

- Adapting C^3RS
- Confidentiality fears
- Cost and efficiency of running C^3RS
- Disputes/Scope Conflicts
- Economy
- Funding
- Labor-management relationships
- Lack of understanding of C^3RS
- Maintaining interest
- National model
- Optimism about C^3RS
- Personnel turnover
- Planning industry model
- Post-pilot
- Program ending
- Public image of C^3RS
- Risk Reduction rule
- Skepticism about C^3RS
- Stovepipes
- Sustainability

Appendix E. Example Survey Cover Letter

Confidential
Close Call
Reporting System

C^{3}RS Lessons Learned Survey

As you know, there is a joint effort by the FRA, UP labor (BLET, SMART TD), and UP management to test a safety improvement process known as the Confidential Close Call Reporting system (C^{3}RS) here at North Platte. If C^{3}RS works, the intention is to invest the resources needed to implement C^{3}RS across the railroad industry. But will it work? Will the investment be worth the effort? To find out, a Lesson Learned Team (Evaluation Team) was organized by the FRA to assess the impact of C^{3}RS on safety and safety culture. The assessment conducted by the Evaluation Team will provide both UP and the FRA with valuable information on C^{3}RS. The Evaluation Team is comprised of the Volpe Center, which is a US Department of Transportation (DOT) research center, Jacobs and Fulcrum Corporation, which are companies that support evaluation of safety initiatives, and the Bureau of Transportation Statistics (BTS), which is a statistical agency in DOT that supports data collection and data analysis.

You will see that the survey does NOT ask for your name. Your anonymity is important to the Evaluation Team. To further protect anonymity, the completed surveys will be sent directly to the BTS. Federal law 107-347 and the BTS Confidentiality Statute (49 U.S.C. 111(k)) gives the BTS the right and the obligation to protect data. By law, BTS will protect the identity of any survey respondent. BTS will not release any survey data collected from individual employees to FRA or any other public or private entity, including UP management. Any data and information collected through this survey will be use by the Evaluation Team for statistical purposes only and summary results will be published in a lessons learned report. The final lessons learned report will be available to all employees at this site. Further guidelines that will be used include:

- Summarized results will be given to the PRT, the C^{3}RS steering committee and selected others.
- FRA will use the findings presented in the final report to deepen its understanding of lessons learned from the C^{3}RS project.
- The lessons learned will be shared with the railroad industry.

What we are asking you to do

- Complete the attached survey, seal it in the envelope provided, and give it to the person conducting your Safety Meeting.
- Use a pencil to mark the responses that best match your opinion.
- The survey looks long, but testing has shown that it takes only about twenty minutes to complete. Please give us those twenty minutes of your time.
- A 100% response rate is important to us. If you know someone who is absent, please encourage him or her to complete the survey. The person handing you this survey will have instructions as to how absent people can get a copy of the survey to complete.
- If you have already filled out this survey and you receive a second copy, please do not fill it out a second time.

Demetra Collia at the Bureau of Transportation Statistics is the survey coordinator for the Lesson Learned Team. If you have any questions about the survey, please call her at: XXX-XXX-XXXX, or send her email at: name @bts.gov

Thank you for your assistance.

Labor Representative Labor Representative Management Representative

Appendix F. Detailed Implementation Assessment Rubric for Each Site

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
FRA Responsibility				
Funding Extent to which FRA budget allocations support C^3RS, in recruiting new railroads, program management, implementation team training, and payment to Third-Party	*Rating = Very Good (4)* Provided adequate and consistent funding for C^3RS Third-Party, C^3RS Implementation Team (from RRS RRP and HPD and Volpe), and software and training materials, the Evaluation Team (funded by RD&T)	*Rating = Very Good (4)* Provided adequate and consistent funding for C^3RS Third-Party, C^3RS Implementation Team (from RRS RRP and HPD and Volpe), and software and training materials, the Evaluation Team (funded by RD&T)	*Rating = Very Good (4)* Provided adequate and consistent funding for C^3RS Third-Party, C^3RS Implementation Team (from RRS RRP and HPD and Volpe), and software and training materials, the Evaluation Team (funded by RD&T)	*Rating = Very Good (4)* Provided adequate and consistent funding for C^3RS Third-Party, C^3RS Implementation Team (from RRS RRP and HPD and Volpe), and software and training materials, the Evaluation Team (funded by RD&T)

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
FRA- Granted Waivers Adequacy of waivers that legally permit participating railroads to refrain from taking disciplinary action in return for employees submitting C^{3}RS reports; and protection for railroads from fines imposed by FRA in the event that an employee was not disciplined	*Rating = Fair (2)* Provided waivers as needed during the pilot. After the pilot, the site requested that FRA continue to provide waivers without continuing in the FRA C^{3}RS program. FRA declined this request and the site withdrew	*Rating = Good (3)* Provided waivers as needed. Due to novelty of legal arrangement, process was slower than ideal	*Rating = Good (3)* Provided waivers as needed. Due to novelty of legal arrangement, process was slower than ideal	*Rating = Good (3)* Provided waivers as needed. Due to novelty of legal arrangement, process was slower than ideal
Perceived Neutrality of FRA Extent to which railroad labor and management perceive FRA to be unbiased with respect to disputes about C^{3}RS's protections	*Rating = Good (3)* Perceived as unbiased	*Rating = Good (3)* Perceived as unbiased	*Rating = Good (3)* Perceived as unbiased	*Rating = Poor (1)* Management perceived FRA as biased during disputes, favoring labor

Assistance from Implementation Team Extent to which the C^{3}RS Implementation Team's efforts are sufficient to accomplish the rollout and operate the PRT including obtaining reports, analyzing cases and developing/implementing corrective actions.	*Rating = Good (3)* The C^{3}RS Implementation Team provided sufficient assistance to get C^{3}RS started at the site and set up a PRT that was functional. Later the FRA ORS established the Human Performance Division to recruit, provide training, and ongoing assistance. The HPD was recognized as providing credible support.	*Rating = Fair (2)* The C^{3}RS Implementation Team provided sufficient assistance to get C^{3}RS started at the site and set up a PRT that was functional. Later the FRA ORS established the Human Performance Division to recruit, provide training, and ongoing assistance. The HPD was recognized as providing credible support. After the pilot, there were concerns about the lack of training available to new PRT members, as membership rotated. After the pilot, senior managers wanted more reports from FRA about what the program has done, including data on	*Rating = Good (3)* The C^{3}RS Implementation Team provided sufficient assistance to get C^{3}RS started at the site and set up a PRT that was functional. Later the FRA ORS established the Human Performance Division to recruit, provide training, and ongoing assistance. The HPD was recognized as providing credible support.	*Rating = Good (3)* The C^{3}RS Implementation Team provided sufficient assistance to get C^{3}RS started at the site and set up a PRT that was functional. Later the FRA ORS established the Human Performance Division to recruit, provide training, and ongoing assistance. The HPD was recognized as providing credible support.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
		types of reports and corrective actions.		
FRA on PRT Extent to which participation in PRT meetings by regional FRA staff supports the analysis of cases and development of corrective actions.	*Rating = Good (3)* FRA provided a member for each meeting, although there was a lot of turnover. Most were highly valued by PRT.	*Rating = Good (3)* FRA provided a member with some turnover. They were highly valued by PRT. After the pilot, opinions varied about how necessary FRA attendance was for every PRT meeting.	*Rating = Good (3)* FRA provided a member with some turnover. Most were highly valued by PRT. FRA PRT members received pressure when they requested travel to PRT meetings but were generally able to attend.	*Rating = Fair (2)* FRA provided a member, and most were valued by the PRT. There was some variation on which individuals were effective.
Third-Party Responsibility				

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Data Collection Effectiveness of data collection by the Third-Party, including the amount and relevance of information collected, and the usefulness of questions on forms and debrief.	*Rating = Good (3)* Reporters generally satisfied with call-back interview process. Analysts were well qualified to ask the right questions.	*Rating = Good (3)* Reporters generally satisfied with call-back interview process. Analysts were well qualified to ask the right questions.	*Rating = Fair (2)* Participants generally satisfied with interview questions. Analysts were well qualified to ask the right questions. But the Third-Party fell behind with the call-back interviews and had long delays after reports were submitted.	*Rating = Good (3)* Reporters generally satisfied with call-back interview process. Analysts were well qualified to ask the right questions.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Detailed C3RS Report Records to PRT</u> Extent to which the PRT is able to understand the C^3RS report records sufficiently to develop an effective corrective action.	*Rating = Good (3)* PRT generally satisfied with the C^3RS report records. Sometimes they asked the Third-Party to add new questions, and they were responsive. Occasional complaints about completeness of data, but it did not get in way of effective analysis.	*Rating = Fair (2)* PRT generally satisfied with the C^3RS report records. Sometimes they asked the Third-Party to add new questions, and they were responsive. Occasional complaints about completeness of data, but it did not get in way of effective analysis. After the pilot, the PRT was somewhat less satisfied with reports from NASA.	*Rating = Fair (2)* PRT complained about the lack of detail. However the PRT did receive many reports with sufficient information to make recommendations for corrective actions.	*Rating = Good (3)* PRT generally satisfied with the Third-Party. Sometimes they asked the Third-Party to add new questions, and they were responsive. Complaints about lack of detail were blamed on lack of detail from the reporters. After the pilot, expressed some dissatisfaction with the call-back interview questions.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Confidentiality Extent to which the Third-Party does not disclose information to the PRT or to the public that might identify the person or who submits a C^3RS report.	*Rating = Poor (1)* There were no breaches of confidentiality. After the pilot, the site (including PRT labor members) expressed their belief that CIPSEA provided more confidentiality than the NASA model. It decided to leave the C^3RS program when FRA switched to NASA.	*Rating = Very Good (4)* There were no breaches of confidentiality. The Third-Party worked to ease concerns and protect data from a request for data from a major news station.	*Rating = Very Good (4)* There were no breaches of confidentiality. The Third-Party worked to ease concerns and protect data from a request for data from a major news station.	*Rating = Very Good (4)* There were no breaches of confidentiality.
Transmission of report records to the PRT Extent to which the Third-Party efficiently transmits C^3RS report records to the PRT in a manner that supports effective action by the PRT.	*Rating = Fair (2)* The Third-Party's confidentially requirements required the inefficient physical mailing back and forth of computers with data.	*Rating = Fair (2)* The Third-Party confidentially requirements required the inefficient physical mailing back and forth of computers with data.	*Rating = Very Good (4)* The Third-Party set up the ability for the PRT to download the information.	*Rating = Fair (2)* The Third-Party confidentially requirements required the inefficient physical mailing back and forth of computers with data.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Carrier Responsibility				
Systemwide Champion Extent to which an influential senior manager is present and active in: 1) protecting and promoting C^3RS within the company, and 2) exerting the authority necessary to implement corrective actions.	*Rating = Good (3)* Enthusiastic support at most senior levels of company for the duration of the pilot. After the pilot, they continued the program for a couple years until FRA decided to move to NASA.	*Rating = Good (3)* Effective support at most senior levels of company. This support varied over time. They were supportive enough to continue the program after the pilot ended.	*Rating = Very Good (4)* Enthusiastic support at most senior levels of company. They expanded the program to include new locations, crafts, and type of reports.	*Rating = Poor (1)* Only had a systemwide champion very early in the pilot, but thereafter senior managers were skeptical of C^3RS.
Managers on the PRT Extent to which management participation on the PRT leads to development and implementation of local corrective actions.	*Rating = Fair (2)* Some managers on the PRT were very actively involved and effective. Others attended PRT meetings less often, were less involved, and had less authority.	*Rating = Good (3)* Effective involvement from managers on the PRT. Met regularly with Support Team. Managers on PRT had other duties, so it was difficult to find time to complete PRT responsibilities.	*Rating = Good (3)* Effective involvement from managers on the PRT. Dedication to C^3RS. Some corrective actions were implemented by PRT members with local managers.	*Rating = Good (3)* Effective involvement from managers on the PRT. They implemented local corrective actions around education.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Local Sponsor(s)</u> Effectiveness of manager(s) whose responsibility is primarily within the boundaries of that part of the carrier that is implementing C^3RS. These managers might not be on the PRT, but are responsible for implementing local corrective actions.	*Rating = Good (3)* Very enthusiastic local support at beginning. Later, as turnover occurred, it varied but was sufficient to keep things going during the demonstration.	*Rating = Good (3)* Very enthusiastic local support at beginning. Later, as turnover occurred, it varied but was adequate.	*Rating = Fair (2)* Varied levels of support in middle management. Some worked with PRT to implemented local corrective actions. Others has little knowledge or interest. Some corrective actions were implemented by local sponsors working with information directly form the PRT.	*Rating = Good (3)* Provided adequate support. They implemented local corrective actions around education.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Cross-functional senior management (Support Team) Effectiveness of the senior management team representing the functions that may need to be involved for the implementation of corrective actions. Includes effectiveness of the team's communication with the PRT.	*Rating = Good (3)* Did not have formal Support Team, but always had contacts at the senior level responsible for implementing corrective actions. Sometimes departments outside Transportation did not support corrective actions.	*Rating = Fair (2)* Early on the Support Team met regularly and has cross-functional members. But during some periods, high turnover on Support Team was disruptive and not many actions were implemented. After the pilot, the Support Team was thought to have improved and been more willing to help. After the pilot, corrective actions were still implemented. There was an impression that only inexpensive actions were implemented, but there were also examples of more significant spending on actions, e.g., more hires in the dispatcher office.	*Rating = Good (3)* The formal Support Team met very infrequently. However, a single consistent senior manager, within the section of the company that was implementing C^3RS, oversaw and coordinated systemwide corrective actions with appropriate corporate contacts as needed. This was effective.	*Rating = Poor (1)* Original Support Team was not effective. Eventually made a new team that implemented some actions. Management seemed willing to implement inexpensive corrective actions, but seemed unwilling to implement expensive physical changes.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Provide resources (for PRT) Extent to which the carrier provides resources for PRT time and travel.	*Rating = Good (3)* Resources provided so PRT was able to meet.	*Rating = Good (3)* Resources provided so PRT was able to meet. Manager estimated this to cost $20-25K per month.	*Rating = Good (3)* Resources provided so PRT was able to meet.	*Rating = Good (3)* Resources provided so PRT was able to meet.
Labor Responsibilities				
Promote Extent to which PRT members promote C^3RS, both formally in union meetings, and informally with peers, and also, the extent to which PRT members are available and willing to answer questions from labor and management.	*Rating = Good (3)* PRT labor members participated in multiple rollouts and promoted the program informally. They promoted reporting that would provide protection from discipline as well as other close calls. They answered questions about the program from labor and management.	*Rating = Good (3)* PRT labor members participated in multiple rollouts and promoted the program informally. They mentioned the benefits of discipline protection and also encouraged other reporting other close call events. They answered questions about the program from labor and management. After the demonstration, labor continued to do this.	*Rating = Fair (2)* PRT labor members participated in multiple rollouts and promoted the program informally. They mentioned the benefits of discipline protection and also encouraged other reporting other close call events. They answered questions about the program from labor and management. A weakness was that not all geographic areas were represented.	*Rating = Poor (1)* A weakness of this site was that one of the participating union's PRT labor members promoted C^3RS as only a "get out of jail" discipline protection program instead of promoting the reporting of many close call events.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Detailed and numerous reports</u> Extent to which eligible reporters provide detailed and numerous C^3RS reports concerning a variety of close calls, and do not just use the program to "get-out-of-jail."	*Rating = Fair (2)* Some reports were sufficient. Some lacked detail. PRT had more reports than they had time to review. After the demonstration, some management believed that C^3RS was being used as "get out of jail" too often.	*Rating = Fair (2)* Some reports were sufficient. Some lacked detail. During some time periods, reporting rates were very low. Some managers expressed some concern about the usefulness of reports received from repeat offenders, i.e., reporters making the same error multiple times. This practice kept them from being able to discipline habitually un-safe people.	*Rating = Fair (2)* Some reports were sufficient. Some lacked detail, partially due to Third-Party actions to protect confidentiality. Labor provision of details improved over time. PRT wished more reports were submitted.	*Rating = Poor (1)* Many reports lacked sufficient detail. PRT seemed frustrated that labor did not report on a greater variety of close calls. The PRT was very concerned that only one person from the crew would report, and they were not receiving sufficient details about the close call. The PRT tried to get two-person reporting required in the IMOU, so each member had to report to receive discipline protection. National labor was against this change and would not agree to the revised IMOU.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Labor on PRT Effectiveness of PRT labor members, e.g., effective participation in meetings, smooth rotation of members on the PRT, "leaving hats off" (i.e., focusing on the goals of C^{3}RS, not the goals of the group they belong to).	*Rating = Good (3)* There was adequate participation and work performed by PRT labor members. One weakness was that management felt too many labor members attended, impacting efficiency.	*Rating = Fair (2)* There was adequate participation and work performed by PRT labor members. However, a weakness was that during some periods, after a union election, there was high turnover on the PRT which was disruptive.	*Rating = Good (3)* There was adequate participation and work performed by PRT labor members.	*Rating = Fair (2)* Most PRT labor members participated effectively in the MCIA analysis, but some members were not focused on C^{3}RS's goals and were only there to defend their members.
Help with corrective actions Extent to which labor helps to implement corrective actions, especially when local action by labor and management can lead to effective change.	*Rating = Very Good (4)* PRT labor members were very involved with implementing corrective actions, especially ones in their area.	*Rating = Good (3)* PRT labor members were involved with implementing corrective actions, especially ones in their area.	*Rating = Good (3)* PRT labor members were somewhat involved with implementing corrective actions, especially ones in their area.	*Rating = Poor (1)* PRT labor members were rarely involved with corrective action implementation.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Shared Responsibilities Includes FRA, Carrier Labor and Management, and Third-Party				
Initial IMOU Existence of an initial signed IMOU between labor, management, and FRA for that site.	*Rating = Very Good (4)* An initial IMOU was signed by all parties	*Rating = Very Good (4)* An initial IMOU was signed by all parties	*Rating = Very Good (4)* An initial IMOU was signed by all parties	*Rating = Very Good (4)* An initial IMOU was signed by all parties
IMOU renegotiations Effectiveness of IMOU re-negotiation processes as needed	*Rating = Poor (1)* No issues requiring renegotiation during the pilot. A critical weakness after the pilot was that they were not able to negotiate a new IMOU to move to NASA and decided to leave the C^3RS program. PRT labor members and management had reservations about NASA	*Rating = Good (3)* Negotiated a new IMOU to move to NASA after the pilot. A minor weakness after the pilot was some need for additional clarity concerning the intent and meaning of certain provisions in the IMOU	*Rating = Very Good (4)* Successfully negotiated expansions and their new IMOUs	*Rating = Poor (1)* PRT labor members and management, Support Team, and Senior Management agreed on issues and tried to modify the IMOU near the end of the pilot. But national labor and FRA could not come to an agreement with the PRT/Support Team/Senior Management

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Confidentiality</u> Extent to which all stakeholders have a common understanding of the terms of C^3RS confidentiality, and the degree to which that confidentiality is protected.	*Rating = Fair (2)* They took confidentiality seriously and maintained trust in the program. A weakness was that many concerns about confidentiality led to unnecessary restrictions in communication about program accomplishments. After pilot, the site (including PRT labor members) was worried about NASA's legal protection and wished that BTS was still involved. There were misconceptions about what information NASA would release.	*Rating = Good (3)* They took confidentiality seriously and maintained trust in the program.	*Rating = Good (3)* They took confidentiality seriously and maintained trust in the program.	*Rating = Good (3)* They took confidentiality seriously.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Training Adequacy of training for PRT concerning the provisions of C^3RS, MCIA process, usage of tools, and roles and responsibilities.	*Rating = Good (3)* Initial training for PRT deemed adequate	*Rating = Good (3)* Initial training for PRT deemed adequate	*Rating = Good (3)* Initial training for PRT deemed adequate	*Rating = Good (3)* Initial training for PRT deemed adequate
Communication between C^3RS Internal Stakeholders Effectiveness of communication between the PRT and the Support Team.	*Rating = Fair (2)* The PRT and Senior managers had varying levels of communication. Sometimes senior managers attended PRT meetings to discuss corrective actions, and the PRT greatly appreciated it. Other times, feedback was slower.	*Rating = Good (3)* The PRT and Support Team met face to face regularly. There were some minor communication issues between them.	*Rating = Fair (2)* PRT wanted more communication from Support Team about the reasons why some corrective actions were rejected. The support team provided feedback on the corrective action tracking documents back to the PRT, but it was not frequent.	*Rating = Fair (2)* Initial Support Team did not respond to requests for corrective actions in a timely manner. Late in the pilot, a second Support Team was more involved and communicated better.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Communication to Workforce and Management External to C³RS</u> Effectiveness of outreach to workforce and management concerning the intent, usage, and achievements of C³RS.	*Rating = Fair (2)* Workforce and managers needed more information about C³RS, especially about achievements of C³RS. PRT conducted occasional rollouts to tell workforce about C³RS.	*Rating = Fair (2)* Workforce needed more information about C³RS, especially about achievements of C³RS. After the pilot, the PRT realized that not everyone in the workforce understood C³RS, especially newer employees, so another rollout was planned.	*Rating = Fair (2)* Workforce needed more information about C³RS, especially about achievements of C³RS. Middle managers also needed more information about C³RS.	*Rating = Fair (2)* Corrective actions often were informational posters for employees, so they could see what C³RS was accomplishing. Some PRT labor members communicated an incorrect intent of C³RS to the workforce, emphasizing discipline protection over providing detailed reports.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Data Analysis Extent to which the PRT's data analysis is able to reveal root and contributing causes, trends, cases that should be dealt with as a group of related issues, and safety priorities for action.	*Rating = Fair (2)* Management wanted more data analysis from PRT (trends, risk analysis, prioritization). Management agreed to provide an expert to help, but the resource was not provided.	*Rating = Fair (2)* Support Team wanted more data analysis from PRT (frequencies and costs). Late in the pilot, management provided an expert to help with data analysis which was productive.	*Rating = Good (3)* PRT bundled related C^3RS reports when creating and recommending corrective actions and looked at frequencies every month. Due to limitations with the MCIA tool, the PRT developed their own documents to structure their analysis.	*Rating = Fair (2)* PRT and managers wanted more data analysis, but many report records lacked sufficient details.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Process Efficiency Extent to which stakeholders work together to make C^3RS processes more efficient.	*Rating = Fair (2)* Management was concerned about PRT cost and efficiency. The MCIA tool (provided by the Implementation Team) was very cumbersome. Each case was analyzed one at a time. The PRT was viewed as too large by management. Later in the pilot, the PRT made some efficiency improvements, including a sub-team to prioritize cases.	*Rating = Fair (2)* PRT made efficiency improvements, including a sub-team to prioritize cases. After the pilot, the PRT had a backlog of 200 cases, indicating more improvements in efficiency could be made.	*Rating = Fair (2)* There was some frustration from PRT managers about time utilization. The PRT decided to cut the frequency of meetings to save costs.	*Rating = Fair (2)* PRT did not focus on efficiency improvements, but management did not indicate that this was a major area of concern. There were complaints from managers about the Third-Party laptops arriving late.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Ability to Implement Corrective Actions				
Cooperation Extent to which relevant parties are able to cooperatively develop and implement corrective actions. "Relevant parties" may include, as needed, PRT labor and management, local management, the Support Team, and senior management.	*Rating = Good (3)* In general the PRT and local management worked together and implemented actions. Cooperation varied with the local managers involved.	*Rating = Good (3)* In general the PRT labor and management worked together and implemented actions. After the pilot, they continued to cooperate in developing corrective actions.	*Rating = Good (3)* In general the PRT labor and management worked together and got local actions implemented actions. A senior manager cooperated with the PRT on systemic issues, however communication was sometimes infrequent. System corrective action recommendations that were rejected were not explained to PRT and led to some PRT discouragement.	*Rating = Poor (1)* There was little cooperation between PRT labor members and management to implement corrective actions; PRT labor members did not help managers to implement actions. After pilot ended, former PRT labor members missed ability to come together and sit in room with management to discuss rules and actions. Management believed that not all sides "left their hats off."

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Tracking Extent to which the participants track and update the status of corrective actions.	*Rating = Fair (2)* Minimal electronic tracking tools, but they had paper-based tracking. There was an attempt to incorporate it into the MCIA tool, but it was not successful at the time.	*Rating = Good (3)* A detailed tracking spreadsheet was developed and kept updated on a regular basis.	*Rating = Fair (2)* Local actions not well tracked or written down. Senior manager occasionally updated text files about larger action and mailed them to all project participants.	*Rating = Poor (1)* A tracking spreadsheet was made but not updated until years later. Several years in, the second Support Team made a large effort to go through all the recommendations and update the decisions and status. Many actions not written down.
Accountability Extent to which specific people are clearly assigned responsibility for implementing each corrective action and held accountable.	*Rating = Fair (2)* Accountability varied over time depending on the superintendent. Senior managers' systemic corrective action accountability got better over time.	*Rating = Good (3)* Responsibility for each action was assigned to a specific person and tracked.	*Rating = Good (3)* The status of system wide corrective actions was communicated to PRT in writing, but there were delays in communication. Senior manager held himself personally accountable.	*Rating = Poor (1)* Little accountability for most of the lifecycle of pilot, decisions made a long time after recommendations. The second Support Team was adequate and did a lot of work assigning corrective actions.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Effective Dispute Resolution				
Come to agreement Extent to which participants in a dispute are able in a timely manner to either resolve the dispute, or agree to "agree to disagree."	*Rating = Very Good (4)* Minor disputes during pilot with no large impact because the resolution was consistent with the spirit of the program, protecting the reporter. Trust was not undermined.	*Rating = Fair (2)* Minor disputes during pilot with no large impact. After the pilot, had a larger dispute concerning a person who had multiple violations and only turned in one C^3RS report. He received protection for only one violation as was specified in the IMOU, but the some PRT labor members were not satisfied with the resolution.	*Rating = Very Good (4)* No major disputes	*Rating = Poor (1)* Two major disputes resulted in discipline. Some PRT members were very displeased with the outcome.
"Move on" after Extent to which the PRT continues to function after a dispute is negotiated.	*Rating = Good (3)* The PRT continued to function after disputes were resolved.	*Rating = Good (3)* After the pilot, one union threatened to withdraw from C3RS due to the outcome of a dispute. However, they eventually decided there were enough benefits to continue with C^3RS.	*Rating = Very Good (4)* No major disputes	*Rating = Poor (1)* Displeasure with outcome of dispute and broken trust. One person removed from PRT.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Perceived Value				
Safety Extent to which C^3RS was perceived as contributing to safety improvements, regardless of whether or not there was empirical evidence to justify such a claim.	*Rating = Good (3)* During the pilot, a variety of opinions existed about the impact on safety. At the end of the pilot, participants agreed that derailments were reduced and C^3RS played a part along with other safety initiatives. After the pilot, senior management was not sure that C^3RS was continuing to provide enough benefit to be worth the cost.	*Rating = Good (3)* Midway through the pilot, some managers had access to data that said things were getting better, but not everyone was aware. Later they saw and agreed that safety (derailments and injuries) improved during high usage of C^3RS. After the pilot, there was a consistent view that C^3RS contributed to improvements in safety.	*Rating = Good (3)* Most people were not aware of improvements. They did not have a consistent picture of measureable change. However participants did agree when the Evaluation Team showed them improvements during the final briefing.	*Rating = Fair (2)* Participants did not observe any improvements in safety metrics.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
Safety Culture Extent to which C^3RS was perceived as contributing to safety culture improvements, regardless of whether or not there was empirical evidence to justify such a claim.	*Rating = Good (3)* Perceptions of improved safety culture and employee engagement by labor and management.	*Rating = Good (3)* Some perceptions of improved safety culture. Stakeholders noticed increase supervisor-labor communication, especially about run-through switches.	*Rating = Good (3)* Some perceptions of improved safety culture.	*Rating = Fair (2)* Some interviewees indicated a decline in safety culture. After the pilot, some people were positively impacted. Some labor more aware of safety issues and willing to discuss issues with managers. Others did not change.
Corrective Actions Extent to which corrective actions that are implemented are perceived to be effective, regardless of whether empirical data exists for their effectiveness.	*Rating = Good (3)* Many actions implemented, people believe they were valuable. There was still a desire to do more.	*Rating = Good (3)* Many actions implemented, people believe they were valuable. There was still a desire to do more.	*Rating = Good (3)* People believe that there were some good actions implemented. In general, most wished there were more.	*Rating = Fair (2)* Participants said some actions were accomplished, but they were not satisfied with the number and effectiveness of corrective actions. After pilot, said that C^3RS revealed short comings in training program that would not have known without C^3RS.

Fishbone	Site 1 Assessment	Site 2 Assessment	Site 3 Assessment	Site 4 Assessment
<u>Cost Savings</u> Extent to which C^3RS and its corrective actions resulted in perceived cost savings (or its time equivalent), regardless of whether or not there was empirical evidence to justify such a claim.	*Rating = Good (3)* Interviewee said that discipline hearings were drastically reduced.	*Rating = Good (3)* After the pilot, PRT reported that reminder signs at end of electric track saved $60K a year in repair costs for pantographs.	*Rating = Good (3)* Interviewees said that discipline hearings were reduced and provided cost savings data.	*Rating = Fair (2)* Participants did not observe cost savings.

Appendix H. Example Corrective Actions

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Excess Speed	1	Employees were having difficulties interpreting and following the track warrant paperwork related to speed restrictions because the formatting of the text required the conductor and/or engineer to read it in a nonstandard manner. At times it required reading right to left instead of left to right and bottom to top instead of top to bottom.	The corrective action implemented made the track warrant easier to read by using standard left to right and top to bottom formatting. To execute this corrective action, corporate software was edited. The improved track warrant format was implemented centrally.
Excess Speed	1	The crew would sometimes be given instructions to change speed many times within a short distance due to maintenance occurring on the track. Engineers could speed up between the areas designated as "slow" due to the track maintenance, but then would have to slow down again for the next area where track was being repaired. This created speeding up and slowing down and sometimes confusing paperwork for the crews which was difficult to follow.	The corrective action was to consolidate the areas that were considered "slow" within 10 miles of each other to reduce the number of required changes in speed. This change involved more than just transportation managers; it required track maintenance managers, too. Due to different reporting chains of command in track maintenance and transportation a high level of senior management involvement was needed. Interviewees indicated that it was difficult to maintain the implementation of the slow order consolidations over time.
Excess Speed	1	Crews forgetting about weather related speed restriction.	Plastic hang tags that were designed to hang in the cab to be a reminder for crews. One tag had a reminder for cold weather restrictions on one side and hot weather on the other side.
Excess Speed	1	Crews forgetting about which slow orders applied, men on track.	Another tag reminded the crews what types of instruction were applicable to their route: Form A slow orders or Form B men around the track.
Excess Speed	1	Crews making mistakes concerning the train speed.	Enhancements to simulator training based on learnings from C^3RS.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Derailment	1	There was an area of the yard that was not visible to the foreman. When he threw a cross-over switch, there was a risk that a train was still traveling in that area and could be derailed by a switch aligned in the direction opposite of its travel.	The corrective action that was implemented placed a camera in the affected area, to assist the foreman in seeing equipment placement prior to throwing the cross-over switch.
Derailment	1	The mechanical department installs derail devices on the track on both ends of a railroad car while they are performing maintenance to prevent the car from moving and harming the worker. Sometimes employees forget to remove the derails, and when someone moves the car, there can be a derailment.	The C^3RS corrective action was to implement a policy to make sure that the derails were removed after the maintenance job was complete.
Derailment / Run-Through Switch	1	If a crew sitting in the locomotive cab cannot tell how a switch is aligned, then the train can run through it in the wrong direction, damaging the switch. Subsequent trains that travel through that damaged switch may have a derailment.	There were several corrective actions implemented to improve the visual cues for the status of switches, e.g., reflective switch targets.
Run-Through Switch	1	At a specific switch, the target was difficult to see.	Installed larger switch targets.
Run-Through Switch	1	Run-through switches in the yard.	Added content to training classes based on learning from C^3RS.
General	1	Crews unaware of important safety information when starting a new shift.	A hang tag with reminder to turn on headlights on one side and a departure checklist on the other side. This was a system level corrective action, as tags were made available to other locations outside of the demonstration pilot limits.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Excess Speed	2	Portions of track have regular speed limits and also temporary areas of speed restrictions based on things such a as track condition. The train crews receive paperwork about the speed restrictions that need to be followed. The speed restrictions were listed on multiple forms creating disorganized and difficult to follow instructions.	Changes to paperwork to keep the daily speed restrictions organized and on one page.
Excess Speed	2	Difficult to hold all of the paperwork in the cab.	They also made a design change and installed a clip in the cab to keep paperwork in sightline.
Excess Speed	2	Crews would be traveling at a restricted speed, then have a station stop. After they resumed, they would forget about the restricted speed.	They installed reminder signs about restricted speeds after station stops.
Derailment / Run-Through Switch	2	There were communication issues between dispatchers and yardmasters; for example, when one crew was trying to enter the yard, the switch would be lined for them. If another crew was exiting the yard at the same time, their reverse move through that same switch would cause them to run through it in the wrong direction, damaging the switch. Damaged switches can cause the next train to have a derailment.	Site 2 implemented several corrective actions related to communication in the yard. For example, a “squawk box” was installed to provide a direct line of communication between the yardmaster and dispatching.
Run-Through Switch	2	If a crew sitting in the locomotive cab cannot tell how a switch is aligned, then the train can run through it in the wrong direction, damaging the switch. Future trains that travel through that damaged switch may have a derailment.	Lighting improvements in some yards, to better allow crews to see switches.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Run-Through Switch	2	Several switches had caused issues over time due to their proximity to another switch, making it difficult to see which switch was being controlled.	Painted track switches to enhance visibility.
Run-Through Switch	2	Train crews that were inexperienced in outbound reverse movements in train yards were at risk for making mistakes and damaging switches.	A team consisting of PRT members from labor, management, and FRA went into the field and trained all the crews how to make outbound reverse movements in train yards.
Collision	2	Roadway worker groups may be unaware of other groups working in their same limits.	Improvements in communications between roadway worker groups
General	2	Crews unaware of important safety information when starting a new shift.	Improvements to job safety briefings.
Excess Speed	3	Dispatchers were making errors labeling speed restrictions because they were changing back and forth between yard and road dispatching in a single shift.	The correction action was a change in dispatching office assignments, so people did not have to change types of work during a single shift; for example, one employee would do only yard dispatching or only road dispatching for their shift.
Excess Speed	3	Dispatchers were making errors labeling speed restrictions because they were changing back and forth between yard and road dispatching in a single shift.	A second corrective action created written instructions about how to apply restrictive labels.
Excess Speed	3	Speeding due to misreading of TSRB, improper job briefing.	TSRB's format modified: 1) clarity in font, characters and delineation, 2) some line segments rearranged for more logical progression.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Derailment	3	In one location, some electronic blue lights were controlled by dispatchers who could not see the track. Incidents occurred when a derail went up while trains were still moving.	The corrective action was to add a yardmaster to the process to confirm that the crew has stopped moving before the blue flag went up.
Derailment	3	Crew making mistakes and going on the wrong track and striking a derail.	Rules training on switch position and shoving.
Derailment / Run-Through Switch	3	Miscommunication between yard control and yard crews. Unclear permission to line up. Resulted in switch run through.	Improved radio procedures reduced radio traffic and radio training module distributed.
Derailment / Run-Through Switch	3	Conductors working in yard they have not worked in a while.	If have not worked yard in 6 months, retraining is required.
Collision	3	There was a risk in the yards of trains hitting all terrain vehicles crossing over the tracks.	The first corrective action was a Safety Bulletin designed to increase awareness about crossing tracks safely. This bulletin was sent out in a mass email and tacked on bulletin boards. It reminded employees to look both ways and not rush in front of moving equipment.
Collision	3	There was a risk in the yards of trains hitting contract vehicles crossing over the tracks.	Another corrective action added a standard contract requirement for contract employees to not use cell phones while driving around the yard.
Collision	3	There was a risk in the yards of trains hitting contract vehicles crossing over the tracks.	Carts were also labeled in some yards, to allow easy identification if someone was not driving safely and thus encourage safer behavior.
Collision	3	There was a risk in the yards of trains hitting employees crossing over the tracks. Employees walked over the tracks when going to work from the employee parking lot.	In one location, barriers were put up between the employee parking lot and a nearby track to prevent people from walking over the track.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Collision	3	Garbage dump trucks were fouling the track when emptying dumpsters. This put them at risk to be hit by a train.	In another, a dumpster was moved to prevent trucks emptying dumpsters from fouling the track.
Blue flag	3	Site 3 uses blue flags or lights to indicate that maintenance is working on a track. C^3RS reports revealed that sometimes a maintenance crew finished work, but did not remove blue flag protection. The train crew was told that track was available, but they arrived and found a blue flag still up and derails in derailing position. This situation caused delays and confusion, as the train crew is not allowed to remove the blue flag themselves and must call for clarification and assistance. Also, there were occasional reports of maintenance working after blue flag protection had been removed.	The corrective action was to revise the standard maintenance procedure to include instructions about taking down blue flags, which was communicated to all maintenance employees company-wide.
Blue flag	3	Further C^3RS reports revealed that there were issues with blue flag signals having missing lights, faded colors, peeling paint, etc.	The standard maintenance procedure underwent a second update to include instructions about blue signal maintenance.
Excess Speed	4	Many different speed restrictions within a short distance.	Some permanent and temporary speed restrictions were combined to make them easier to follow.
Excess Speed	4	Crew would sometimes leave old paperwork in the cab after their shift, and the next crew would get confused about which paperwork applied.	A poster to remind employees to remove old paperwork at the end of their shift.
Excess Speed	4	Crews would forget about some speed restrictions during a trip.	A poster to notify crews to review paperwork before initial train movement.
Excess Speed	4	Crews would get confused about speed restrictions.	Held an informational meeting to review speed restrictions.

Analyzed-Case Category	Site	Problem Description	Corrective Action Implemented
Excess Speed	4	Speed signs that were difficult to follow.	Modified speed signs.
Run-Through Switch	4	Crews were sometimes struggling with remote control switches.	The C^3RS PRT created a flyer about activating a specific remote-control switch. The flyer included diagrams to prevent run-through switches in a location that was known to have had frequent problems.