U.S. Department
of Transportation
**Federal Railroad
Administration**

Human Reliability Analysis in Support of Risk Assessment for Positive Train Control

Office of Research
and Development
Washington, DC 20590

U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe National Transportation Systems Center
Cambridge, MA 02142

Human Factors in Railroad Operations

DOT/FRA/ORD-03/15

Final Report
June 2003

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

Notice

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE June 2003	3. REPORT TYPE AND DATES COVERED Final Report March 2001-December 2001

4. TITLE AND SUBTITLE Human Reliability Analysis in Support of Risk Assessment for Positive Train Control	5. FUNDING NUMBERS R3103/RR304
6. AUTHOR(S) John Wreathall, Emilie Roth, Dennis Bley, and Jordan Multer	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Department of Transportation Research and Special Programs Administration John A. Volpe National Transportation Systems Center Cambridge, MA 02142-1093	8. PERFORMING ORGANIZATION REPORT NUMBER DOT-VNTSC-FRA-03-03
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Department of Transportation Federal Railroad Administration Office of Research and Development 1120 Vermont Avenue, NW Mail Stop 20 Washington, DC. 20590	10. SPONSORING/MONITORING AGENCY REPORT NUMBER DOT/FRA/ORD-03/15
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161. This document is also available on the FRA web site at www.fra.dot.gov.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report describes an approach to evaluating the reliability of human actions that are modeled in a probabilistic risk assessment (PRA) of train control operations. This approach to human reliability analysis (HRA) has been applied in the case of a safety evaluation of the Communications-Based Train Management (CBTM) System being tested by CSXT Transportation, Inc. (CSXT). This report describes the overall approach to the HRA and its trial application to the CBTM evaluation.

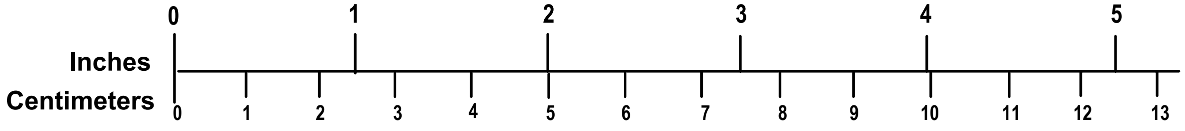
14. SUBJECT TERMS cognitive task analysis, communications, decision-making, human factors, human reliability analysis (HRA), positive train control (PTC), probabilistic risk assessment (PRA), railroad operations	15. NUMBER OF PAGES 140
	16. PRICE CODE

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

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 CONVERSION

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

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 report describes an approach to evaluating the reliability of human actions that are modeled in a probabilistic risk assessment (PRA) of train control operations. It describes an approach to human reliability analysis (HRA) and its application to a test case called Communications Based Train Management (CBTM). The Federal Railroad Administration's Office of Research and Development funded this research effort.

Many people and organizations made the completion of this task possible. We would like to thank the members of the RSAC PTC task force for the guidance and support they provided us throughout the HRA quantification project. We would particularly like to thank Ms Denise Lyle of CSXT for all her support. Denise facilitated our visits to the yard in Spartanburg, South Carolina to conduct interviews and observations of CSXT Locomotive engineers and conductors, and the CSXT Dispatch Center in Jacksonville, Florida, to interview and observe CSXT railroad dispatchers. She also responded kindly and promptly to our many requests for information. We would also like to thank Mr. Tim DePaepe, of the Brotherhood of Railway Signalmen (BRS), Mr. Bob Harvey of the Brotherhood of Locomotive engineers (BLE) and Dr. Fred Gamst of the University of Massachusetts for their thoughtful inputs and their careful review and comment on early drafts of the results of the two site visits and quantification materials.

We would also like to acknowledge the support we obtained from the local CSXT management and labor representatives in Spartanburg, South Carolina and Jacksonville, Florida. In particular we would like to acknowledge the support of Mr. Franky Allan, CSXT train master in Spartanburg, South Carolina and Mr. Terry Davis, local BLE representative in Augusta, Georgia, who facilitated our site visit and helped to recruit CSXT locomotive engineers and conductors to participate in the focus groups in Spartanburg, South Carolina. We would also like to acknowledge the support of Mr. Crawford Boggs, local labor representative for the dispatchers in Jacksonville, Florida, who helped to recruit CSXT dispatchers to interview and observe. We also owe a debt of gratitude to Mr. C. Brock Lucas and Mr. John Campbell of CSXT facility in Jacksonville, Florida for sharing their facility with us and enabling us to observe dispatcher operations over several days and talk with their training staff. Most importantly, we would like to thank all the individuals who participated in the interviews and focus groups held in Spartanburg, South Carolina and Jacksonville, Florida. They provided invaluable knowledge about local operations and CBTM.

We are also grateful for all the support we received in organizing the Human Factors Quantification Workshop. We would particularly like to thank the labor representatives on the PTC RSAC task force for their active support in identifying and recruiting local and national labor representatives to participate in the workshop. Their active participation was critical to the conduct of the workshop.

Most importantly we would like to thank all the individuals who participated in the two-day Human Factors Quantification workshop in Greenville, South Carolina: Crawford Boggs, Sherry Borener, Calvin Cummins, Terry Davis, Tim DePaepe, Manuel Galdo, Fred Gamst, Louise Gunderson, Bob Harvey, Jerry Hicks, Tim Hudson, Rick Inclima, Ron Lindsay, Denise Lyle, Robert McCown, Richard McCord, Ashleigh Merritt, Wayne Minger, Howard Moody, Bob Ralph, Jack Ramsey, Thomas Raslear, James Stem, Tom Sullivan, William Thompson, and Terry Tse.

We would like to thank Dr. Ted Giras, Dr. Lori Kaufman, and Mr. Marc Monfalcone for their help in explaining how the ASCAP model works and indicating how they could incorporate the results of our human reliability analysis.

We are also indebted to Sherry Borener, of the Volpe Center, for helping us understand the contents and limitations of the FRA databases.

We would also like to offer Wayne Minger special thanks for the advice he offered for moving the study forward in a positive and productive way. He helped translate the language of engineering into human factors terms.

Finally, we would like to thank Dr. Thomas Raslear, Mr. Robert McCown, Mr. William Goodman, and Mr. Grady Cothen of the FRA for their support in this effort.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Overview

The railroad industry is developing a new generation of processor-based signal and train control systems to improve safety and enhance operations. To meet the challenge of enabling railroads to adopt new signal processor based technology while reducing risk, the Federal Railroad Administration published the Notice of Proposed Rule Making (NPRM), "Standards for Development and Use of Processor-Based Signal and Train Control Systems; Proposed Rule," (Department of Transportation, 2001). The NPRM proposes that probability-based risk analyses (PRAs) be used as part of a performance-based standard to evaluate the risk associated with the introduction of new systems.

Humans play a very important role in ensuring safety with the current train control systems. Actions include stopping trains when reaching the ends of approved track occupancy (either signal- or block authority-based), keeping train speeds within approved limits, maintaining separation from roadway workers and work locations, and taking actions when things generally "go wrong." The NPRM specifically identifies the need to consider human actions, including their ability to provide "coverage" (i.e., to correct or overcome failures) for the automatic systems.

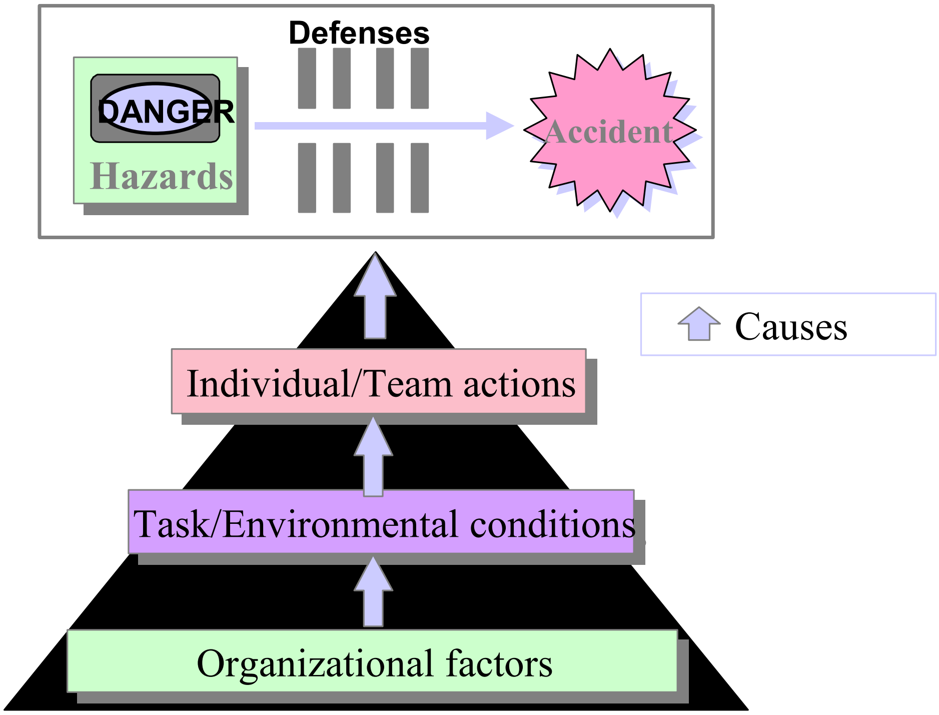
Any meaningful PRA needs to examine human actions (errors, decisions, work-arounds, circumventions, etc.) in a way that accounts for what is known about human performance in technological environments and how human errors can result. This report describes a general human reliability analysis (HRA) methodology for analyzing human performance and estimating the reliability of human actions that can be used in support of PRAs being performed as part of the Product Safety Plan (PSP) submissions to the FRA. In order to exercise and illustrate the HRA approach, it was applied to the safety evaluation of the Communications-Based Train Management (CBTM) System being tested by CSX Transportation, Inc. (CSXT). The report describes the overall approach to the HRA and its trial application to the CBTM evaluation.

The report includes a set of guidelines and recommendations for performing a human reliability analysis to insure that the results will be credible, acceptable to the broad set of stakeholders, meet accepted standards for human reliability analysis, and able to be integrated into probabilistic risk assessments.

It is intended to provide guidance for both organizations that are trying to develop an HRA plan as well as regulatory agencies such as the FRA charged with evaluating an HRA analysis that may be submitted as part of a product safety plan.

Approach for Human Reliability Analysis

The purpose of human reliability analyses is to estimate the likelihood of particular human actions (that may prevent hazardous events) not being taken when needed, or other human actions that may cause hazardous events (by themselves or in combination with other conditions) occurring. Failures to take action to prevent hazardous events, and actions that cause hazardous events, are commonly called "human errors" in HRA. This term does not imply that people are necessarily personally responsible or culpable in some way, just that an action was omitted (or taken) that adversely influenced safety.



(Adapted from *Managing the Risks of Organizational Accidents,* Reason, 1997)

Figure E-1. Relationship of Safety, Human Errors, and Their Influences

Figure E-1 shows a top-level representation of human performance, how human errors can create weaknesses in safety defenses, and how those human errors are conditioned by the environment in which people work. At the very top level, potentially hazardous situations (such as train collisions with other trains or roadway workers and derailments due to overspeeding) are prevented from becoming accidents through defenses being in place. The defenses include the train crew complying with the rulebook of operations, the use of the computer-aided dispatch system (CADS), adhering to speed limits, and the application of fail-safe design principles. Fail-safe design seeks to eliminate the hazardous effects of a failure by having the failure result in non-hazardous consequences.

It is the purpose of the HRA task to estimate the probabilities of human errors that can potentially fail the defenses. However, this estimation needs to take into account the work environment and task conditions under which the work is done, since these can provide an important influence on the likelihood of error. For example, bad weather, long shift times, and high workload all can increase significantly the likelihood of human errors. In turn, work environment and task conditions are often influenced by organizational factors like work rules, duty times, and so on. Therefore, the error estimation process needs to account for these *contributing factors*.

Human reliability analysis employs a set of tools to estimate the likelihood of required human actions being performed when needed. These likelihoods can then be incorporated into the overall risk assessment, so they can be combined with other probabilities, such as those of equipment faults and other hazardous states, to estimate the overall likelihood of hazardous events.

There are four main tasks that need to be performed as part of an HRA. These tasks represent the general process by which human reliability analysis supports probabilistic risk assessment tailored to railroad operations. The details of these steps may vary in each application.

1. *Qualitative Evaluation of Human Factors Issues.* Analyze the impact of the current work environment and new technology on human performance. This task requires study of operating rules, procedures, available data, as well direct observation of the work environment and interviews of individuals involved in the work. The goal is to identify the major sources of human risk and reliability with and without the new system as well as to understand the factors in the current environment that enable errors to be caught and recovered.
2. *Survey of Databases for HRA Sources*. Identify collections of data that may be relevant to the quantification of errors, problems associated with direct application of that data, and ways in which experts in operations can evaluate and adjust that data to the case at hand.
3. *Quantification*. Develop quantitative estimates of the likelihood of the human actions in question. The process for quantification always begins with an evaluation of the relevance of available data to the actions under analysis. The data often provide a broad base for estimation, but almost all databases have limitations and gaps (such as the criteria for events to be recorded) compared with the modeling requirements of the PRA. In many cases an expert estimation process is used to make adjustments for these limitations and gaps.

   One approach is to conduct an expert elicitation workshop that brings together experts in human factors, HRA and PRA and people with extensive experience in railroad operations to examine the available data and agree on plausible quantifications. The operations experts examine the models and assumptions to ensure that they represent the system as it is (or will be) operated. Experts in analysis and operations then jointly examine the available data and agree on adjustments to compensate for known limitations. For many events there will be no relevant tabulated data. In such cases, the workshop facilitators elicit the best available evidence from the experience of the experts in operations, which is then used as a basis for direct estimation of the error probabilities of interest.

   The error probabilities are represented by distributions rather than a single-point estimates so as to explicitly represent the range of uncertainty in the estimate.
4. *Documentation*. To permit review and later understanding of the details of the quantification, all results and processes must be well documented, providing the bases for all estimates.

Section 2 of the report provides a detailed description of the steps involved in these four main tasks. The steps in the HRA process include:

- Identify the specific unsafe actions to be estimated, as defined by the context of the PRA.

- Perform a qualitative human factors analysis to identify the major factors contributing to human risk and reliability.
- Identify the relevant data sources for each action to be modeled.
- Identify the limitations and gaps in each data source as related to the actions being modeled.
- Implement an expert elicitation process to overcome the limitations and gaps in the data sources.
- Synthesize and document the results.
- Perform a review of the results by people familiar with train control operations to make sure the analyses and results are compatible with their experience.

Example Analysis for CBTM Study

In developing the NPRM, the FRA and members of the Railroad Safety Advisory Committee (RSAC) task force charged with developing the rule were concerned with how to assess safety of railroad operations using the new systems; i.e., what is the impact on operating risk. While the proposed rule allows for use of both qualitative and quantitative risk assessment methods, the FRA has supported the development of a quantitative simulation approach called the Axiomatic Safety-Critical Assessment Process (ASCAP) developed by the University of Virginia (Kaufman & Giras, 2000; Monfalcone, Kaufman, & Giras, 2001).

A major objective of the HRA project was to provide a demonstration of the HRA quantification process as input to risk quantification models such as ASCAP. The CSXT CBTM safety case was used to illustrate the methodology. CBTM is a form of train control that provides a warning to the locomotive crew when the train is predicted to exceed the limits of its authority and stops the train if the operator fails to act in time.

The HRA process outlined above was used to estimate human reliability values for input to ASCAP. This involved:

- Estimating human reliability values for the base case: current railroad operations in the territory where CBTM was tested. This CSXT territory was located between Spartanburg, South Carolina and Augusta, Georgia. It was largely “dark territory,” with direct train control (DTC) as the method of operation. The operations analyzed in this study were exclusively DTC.
- Examining the potential impact of CBTM on human performance and human reliability when added to the current DTC operations in the above territory.

The study analyzed the probabilities of specific human errors representing potential contributors to the risks being modeled in the ASCAP study of the CBTM system:

1. Train enters a block without authorization
2. Train exceeds the track speed limit
3. Train enters a preplanned work zone (published in the train bulletin) without authorization

4. Train crosses a misaligned switch

The CBTM system can potentially reduce the likelihood of occurrence of these events; they fall within the set of functions PTC was intended to address. Therefore, the analysis was performed for the base case (current operations without CBTM) and the case when CBTM is operational. Other accident scenarios, such as those involving grade crossings or collisions with “Hi-rail” vehicles used by inspectors were not modeled because they were not affected by the planned use of CBTM and therefore are not part of the ASCAP study. These represent important risks and would be analyzed for new systems that could affect them.

Qualitative Analysis

The qualitative human factors analysis involved two aspects: (1) an analysis of the current work environment to understand the types of errors that can arise and the factors that contribute to those errors; and (2) an examination of the proposed CBTM system, it’s user interface and proposed human-system interaction, to assess its potential impact on human performance and human reliability.

An early prototype of the CBTM system was being tested on the CSXT territory between Spartanburg, South Carolina and Augusta, Georgia. This provided us an opportunity to (1) directly examine its user interface features and observe its operation, and (2) get input from CSXT locomotive engineers and trainers who had familiarity with the prototype CBTM system.

As part of qualitative analysis, two site visits were conducted: a visit to the yard in Spartanburg, South Carolina to interview and observe CSXT locomotive engineers and conductors, as well as to ride a locomotive equipped with the CBTM system; a visit to the CSXT Dispatch Center in Jacksonville, Florida, to interview and observe dispatchers to understand CSXT dispatch operations and the factors that could contribute to dispatcher errors.

The results of the interviews and observations provided the background necessary for structuring the topics covered in the elicitation of expert evidence and estimation of probability distributions that occurred during a Human Factors Quantification Workshop that was conducted as part of the HRA quantification process.

Quantitative Analysis

The primary tasks in the quantitative analysis were the identification of relevant sources of data, specification of their limitations and gaps, and application of an expert elicitation process to compensate for these limitations and gaps.

Two kinds of data are required in HRA studies: information about the numbers of events similar to those being modeled, and information about the number of opportunities for such events so that a probability or frequency of the events can be estimated. Two major sources of data were identified in this study: databases maintained by the FRA, and databases maintained by CSXT. Both sources contain information about the frequencies of events and the opportunities for such events.

While these databases contained relevant information, they exhibited certain limitations and gaps with regard to the events being analyzed. In order to compensate for these limitations, the data needed to be filtered and scaled. To perform these adjustments, a two-day expert elicitation workshop was held on October 29 and 30, 2001, in Greenville, South Carolina. Thirty attendees participated in the workshop including: four railroad representatives and associated consultants;

thirteen workers, union representatives and associated consultants; six FRA representatives and associated consultants; one University of Virginia (ASCAP contractor) representative; and six Volpe Center and associated consultants (including the HRA team).

The formal process for elicitation of expert evidence and estimation of probability distributions is discussed fully in the main report. The final probability estimates for the human error events were computed based on the combination of the databases and expert judgments and generally took the form of probability distributions.

Results

Train-caused Block Boundary Exceedances

This event involved a train entering a block for which it does not have authority because of an error by the train crew. Based upon the available data sources, two paths potentially existed to analyze the likelihood of train-caused block boundary exceedances. One was to use the CSXT disciplinary data that were associated with all CSXT operations, and the second was to focus on the experience within the trial territory (between Spartanburg, South Carolina and Augusta, Georgia).

The CSXT-wide analysis led to an estimate of the exceedance rate to be a distribution with a mean of 3.14 x 10^{-7} events per train-mile. The territory-specific estimate was a distribution having a mean of 5.26 x 10^{-7} per train mile. This difference of a factor of two was considered not significant, given the number of assumptions used to generate them. The ASCAP analysis modeled the exceedance rate per block, not per train-mile, in its estimates. Given that there were 19 blocks along the test territory of 120.5 miles, the average block length was 6.3 miles. Therefore, the mean exceedance rate per block using the CSXT experience was 1.99 x10^{-6} events per block, and using the territory experience was 3.34 x 10^{-6} per block. To select between these two results, their distributions were compared. The comparison is shown in Figure E-2.

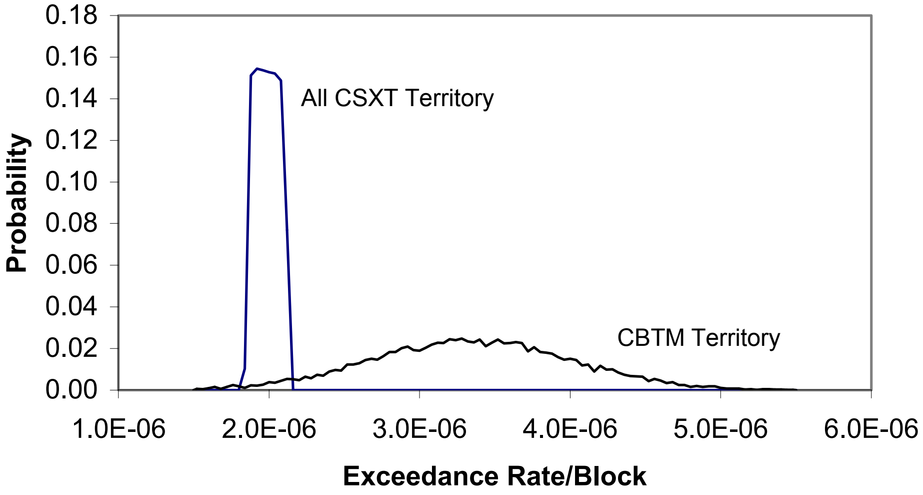


Figure E-2. Comparison of Results for CSXT and Territory Experience

The two distributions overlap, with the territory specific distribution (labeled CBTM territory) extending past the CSXT wide distribution (labeled All CSXT Territory). Based on this comparison, the workshop participants agreed that the CBTM territory result should be used since its mean was slightly more conservative, and its distribution enclosed that of the CSXT-wide analysis. Therefore, the distribution for use in ASCAP for the probability of a train crew to exceed its limit of authority can be approximated by a normal distribution having a mean value of 3.3 x 10^{-6} per block boundary and a standard deviation of 6.8 x 10^{-7}.

A similar quantification process was used to provide probability estimate distribution for each of the other human error events analyzed for the base case. The following results were obtained:

- *Dispatcher-caused Boundary Exceedances*: The mean rate for dispatcher-caused exceedances was 3.5 x 10^{-6} exceedances/block boundary.
- *Overspeeding Events:* The calculated rate for exceedances per restriction was a distribution with a mean of 4.6 x 10^{-6} exceedances per speed restriction.
- *Switches:* Two switching errors were considered: the likelihood of a manual switch being left in the wrong position, and the likelihood of a train running over a mis-positioned switch. The distribution of the likelihood of a switch being in the wrong position at the time a train approaches had a mean value of 1.3 x 10^{-4} per train. Of the 10 manually positioned switches along the length of the route, crews stated that (because of the visibility of the specific switch targets) they would not be able to observe the state of 7 switches when traveling southbound and 6 switches when northbound in sufficient time to stop before running over the switches when traveling at track speed. Of the 3 southbound and 4 northbound switches where the potential existed for stopping, the distribution of the probability of being able to stop in time when traveling at track speed had a mean value of 0.22. If traveling at slow speed (less than 10 mph, as if expecting to enter the siding), the likelihood of failing to stop was considered very low (1 in 10,000).
- *Work Zones*: Data necessary for this event were not available. Workshop attendees suggested using the same fraction as for exceeding DTC block authority

Three conditions were analyzed for the use of the CBTM system:

1. The crew fails to gain control of the locomotive/train following indication of a warning before the penalty brake is applied
2. Train crew over-relies on CBTM (a complacency effect)
3. The train crew enters incorrect consist information into the CBTM system

These three events were selected for quantification based on requirements defined by the PRA, as well as results of the qualitative analyses that were conducted prior to and during the quantification workshop that suggested that these events were situations of potential concern.

The workshop attendees agreed that there was insufficient experience with the CBTM system to confidently project its potential impact on human performance. The local CSX locomotive engineers and conductors indicated that while they had the most experience with CBTM, they have only had the opportunity to operate CBTM equipped trains a couple of times each. Further, the field-tested version of the CBTM prototype was expected to improve substantially prior to actual implementation. Consequently, experience with the CBTM prototype was not expected to be representative of performance of the final production system.

Given the level uncertainty with respect to the likely impact of CBTM on human performance, participants recommended performing sensitivity studies to explore how different assumptions about the impact of CBTM on human reliability would affect the results of the CBTM case.

The results for each of the three individual CBTM issues discussed at the workshop are summarized in the main body of the report. In some cases numeric probability estimates were elicited from the workshop participants. These estimates are presented along with the assumptions that served as a basis for the probability estimates. These probability estimates are recommended as starting points for sensitivity analyses.

Conclusions

The HRA methodology was able to generate reasonable results (i.e., acceptable to the workshop participants) despite the fact that there was no directly applicable database.

The workshop format permitted experts from many different organizations and backgrounds to work together and reach consensus. Uncertainty was expressed through probability distributions that were accepted by the group. The HRA and PRA/ASCAP teams reached agreement that the HRA results were appropriate for use in the PRA.

The approach taken in this study provides one viable way for others to perform HRA studies in support of the FRA's proposed Standards for Development and Use of Processor-Based Signal and Train Control Systems. The lessons learned from performing this example analysis of the CBTM system were documented and provide guidance on avoiding potential pitfalls in future human reliability analyses studies.

Although participants thought the approach worked well, there were several areas of concern:

Biases in data. Data from operational exposure databases or from the experts' opinions has the potential to contain biases that lead to incorrect estimates of probabilities. The approach taken in this study has been to review these databases for potential limitations and biases in the reporting requirements for the databases, review these limitations and biases with the workshop attendees, and make filtering and scaling adjustments based on the inputs of the participants. We recognize that these adjustments represent opinions and the adjusted values may still contain biases. As discussed in the main report, we took steps to limit the potential for significant biases in these opinions, but there is no guarantee that the results are entirely free from bias.

Level of modeling of human error events. The HRA task estimated the likelihood range for the human actions of concern, such as entering a block for which the train has no authority. In contrast the ASCAP simulation modeled human error events at a smaller level of decomposition, explicitly modeling errors in perception and action, and failures to recover ('coverage') from these errors. The rationale for the level of modeling adopted in the HRA study and recommendations for ways to deal with the potential mismatch between the ASCAP and HRA modeling are provided in the main body of the document.

Modeling of future CBTM operations. When the current HRA study took place, the CBTM system was still undergoing field trials, its design was not finalized, and only a limited number of engineers, conductors, and dispatchers had experience with the system. These factors limited our ability to predict the likelihood of errors with confidence. Nevertheless, interviews with engineers and conductors who had experienced the trials of the CBTM system, and discussions held during the expert elicitation workshop enabled identification of potential areas of design and operation that might result in errors or other operational problems. Sensitivity analyses were recommended as a strategy for dealing with the high level of uncertainty associated with the potential impact of CBTM on human performance.

Recommendations for Future Analyses of Rail HRA Studies

The analytical situation that arose in the present study, having some relevant data but with a variety of limitations (not a perfect match for what we want to estimate, with sources that may lead to both under- and over-estimates of frequency) are far from unique to our case. They happen often both in the railroad industry and other industries, and must be addressed explicitly.

The approach we took for combining 'hard data' with expert judgment is a good approach that could be used in other applications. It uses 'hard data' to ground the experts judgments, while using expert judgment to compensate for the known limitations of the existing data.

Guidelines for human factors and human reliability analyses were generated based on the results of this project and are included in Appendix E of this document. The guidelines are intended for organizations developing an HRA plan as well as regulatory agencies such as the FRA charged with evaluating an HRA analysis submitted as part of a product safety plan. Recommendations include:

1. Use an HRA team that includes members experienced in performing human factors studies, human reliability analyses, probabilistic risk assessments, and group facilitation.

2. Model human errors at compatible levels in the PRA and HRA tasks, preferably at the level of available data and experience.

3. Verify that the data sources (databases, expert judgment or a combination) are suitable for the tasks and associated errors being analyzed. Identify gaps or mismatches and utilize expert judgment to leverage the available data while compensating for the known limitations of the data.

4. Conduct qualitative task analyses with people experienced in using the existing systems. Activities should include interviews with workers using the existing systems or the target users of the system (in the case of technologies under development), their trainers and supervisors, so that all levels of experience are included.

5. Utilize expert elicitation methods that take into account known biases and other limitations of expert judgment. Experts should express their opinions in terms of ranges rather than single point values.

6. Solicit input from as broad a range of stakeholders as possible so that the analysis takes into account a wide range of perspectives. Accept quantitative inputs only during the elicitation process, from people with relevant operating experience.

7. Ask the broadest range of stakeholders possible to review the *results* of the analyses to foster support for the results.

1. INTRODUCTION

This report describes an approach to evaluating the reliability of human actions that are modeled in a probabilistic risk assessment (PRA) of train control operations. This approach to human reliability analysis (HRA) has been applied in the case of a safety evaluation of the Communications-Based Train Management (CBTM) System being tested by CSXT Transportation, Inc. (CSXT). This report describes the overall approach to the HRA and its trial application to the CBTM evaluation.

1.1 Use of Risk Assessment for FRA

Historically, the evaluation of train control systems has been design-based. That is, components of a train control system were evaluated based on engineering performance criteria taking into account operability, reliability, and maintainability criteria. With the advent of recent changes in electronic technology, FRA and the railroad industry felt that new and better train control systems might be adopted more quickly using a performance-based approach, assuming that safety could still be assured.

FRA and the industry agreed that performance standards should be based on accident risk assessment and that a quantitative assessment of safety risk associated with any new system should favorably compare against the existing system. Safety or accident risk is defined as the product of the probability of an accident and a measure of the severity or consequences of that accident.

The requirements for performing a quantitative risk assessment are contained in the Federal Railroad Administration's (FRA's) proposed Standard for Development and Use of Processor-Based Signal and Train Control Systems (Department of Transportation, 2001). The proposed rule addressed the development of positive train control (PTC) systems made possible by the introduction of emerging technology in processor-based signal and train control systems. Positive train control systems address three core functions:

- Preventing train-to-train collisions;
- Enforcing speed restrictions and temporary slow orders;
- Providing protection for roadway workers and their equipment.

The complexity of these technologies (communication and information technology) requires additional safety considerations that current safety evaluation methods do not address.

The proposed rule adopted a performance-based approach to enable flexibility in the design and implementation of PTC systems while providing a mechanism to achieve safety goals. The performance standard adopted in the rule requires that the new product or system must not degrade safety below the level of the existing system. To evaluate whether this condition is met requires a risk assessment comparing the new system to the system it will replace.

This proposed rule would require that any railroad wishing to use a processor-based control system (such as a PTC system) to provide more effective or efficient control of train movements must submit a Product Safety Plan (PSP) that includes a quantitative risk assessment that compares the Mean Time to Hazardous Events (MTTHE) for related railroad operations with and without use of the processor-based control system to show that there would be no reduction in safety from implementing the system. The proposed rule also requires that MTTHE values must incorporate the impact of all elements of the system. These elements include human factors as well as the hardware and software components.

While this rule is not final, it is considered very likely that the final rule will contain the same conceptual requirements for performing a quantitative risk assessment as part of the PSP.

In developing the proposed rule, the FRA and members of the Railroad Safety Advisory Committee (RSAC) task force charged with developing the rule were concerned with how to assess risk. Methods for estimating risk vary in complexity from parametric extrapolation of accumulated experience to quantitative modeling (Hollnagel, 1998). While the proposed rule allows for use of both qualitative and quantitative risk assessment methods, the FRA has supported the development of a quantitative modeling approach called the Axiomatic Safety-Critical Assessment Process (ASCAP) developed by the University of Virginia (Kaufman & Giras, 2000; Monfalcone et al., 2001). The ASCAP model considers all types of failures (including human) and is intended to estimate the overall risk—both the probabilities of accidents and the measures of their consequences. ASCAP may be used to determine the comparative risk of the base case vs. an alternative, in this case CBTM.

Train control systems have associated accident risks from non-human failures (i.e., mechanical, electrical, and electronic, materials) as well as human failures. This study focused on:

1) The development of an approach to assess only the human failures in train control systems;

2) The use of that approach to estimate probabilities of human failures on the Spartanburg subdivision of the CSXT railroad under its current train control system (base case);

3) Estimation of likely human failure probabilities under a new and different type of system (CBTM) that overlays on the existing one; and

4) Formatting and defining those human failure probabilities for use in the ASCAP model.

1.2 Role of HRA

Within the scope of the PRA, it is necessary to include human actions and errors that can lead to (or prevent) the hazardous events whose frequencies are to be estimated. Humans play a very important role in ensuring safety with the current train control systems. Actions include stopping trains when reaching the ends of approved track occupancy (either signal- or block authority-based), keeping train speeds within approved limits, maintaining separation from roadway workers and work locations, and taking control when things generally "go wrong." The draft rule specifically identifies the need to consider human actions, including their ability to provide "coverage" (i.e., to correct or overcome failures) for the automatic systems.

Human reliability analysis employs a set of tools to estimate the likelihood of these human actions being performed when needed. These likelihoods can then be incorporated into the overall risk assessment, so they can be combined with other probabilities, such as those of equipment faults and other hazardous states, to estimate the overall likelihood of hazardous events.

Unlike the generally well-documented and accepted methods for estimating hardware failure probabilities, methods for estimating human reliability parameters are not well matured. There exist a wide range of available methods—see the review by Gertman and Blackman (1994), which documents many different approaches. Many more have been developed since that review. However, there is a growing recognition that the most effective methods are those based on failure data for the actual operating experience of the system being modeled and gathered over the widest range of field conditions. However, as with CBTM and other PTC systems for which there is little or no actual operational experience, the preferred HRA methods are those that can combine operating data with modeling or judgment since the operating data by themselves are insufficient or not directly associated with the system being modeled. This is the approach taken in this study.

Since the data are necessarily incomplete (the system not yet being in operation) or only partly relevant to the system being modeled, it is necessary to consider that the sparseness of the data and the judgments needed to supplement the data may introduce uncertainties in the results. In addition, we often do not have complete knowledge of the effects of all the factors that influence the human performance being modeled, another source of uncertainty in the predictions. In order to represent these different sources of uncertainty, we have taken the approach of explicitly representing these uncertainties by calculating distributions rather than providing single-point estimates for the probabilities of human error. More details, together with the overall approach to managing the different sources of uncertainty are presented in Section 2.3.3.

1.3 Purpose of Report

The purpose of this report is two-fold: first, to describe a general HRA estimation process that can be used in support of PRAs being performed as part of the PSP submissions to FRA under the proposed standard described; and second, to present the steps in, and results of, the application of this process in the PRA of the CBTM system being tested by CSXT.

Section 2 of the report describes the principal steps in the HRA process that can be used in other applications. Section 3 presents the principal results of applying the method in the analysis of the CBTM system, with detailed results of the qualitative analysis being presented in Appendix A and B and those of the quantitative analysis in Appendix C. Section 4 of the report presents the lessons learned for future applications of the HRA modeling in future studies, and Section 5 summarizes the recommendations for future studies and the conclusions of this work.

2. APPROACH TO ESTIMATION OF HUMAN RELIABILITY IN TRAIN CONTROL SYSTEM STUDIES

2.1 Overall Approach

The purpose of human reliability analyses is to estimate the likelihood of particular human actions (that may prevent hazardous events) not being taken when needed, or other human actions that may cause hazardous events (by themselves or in combination with other conditions) occurring.

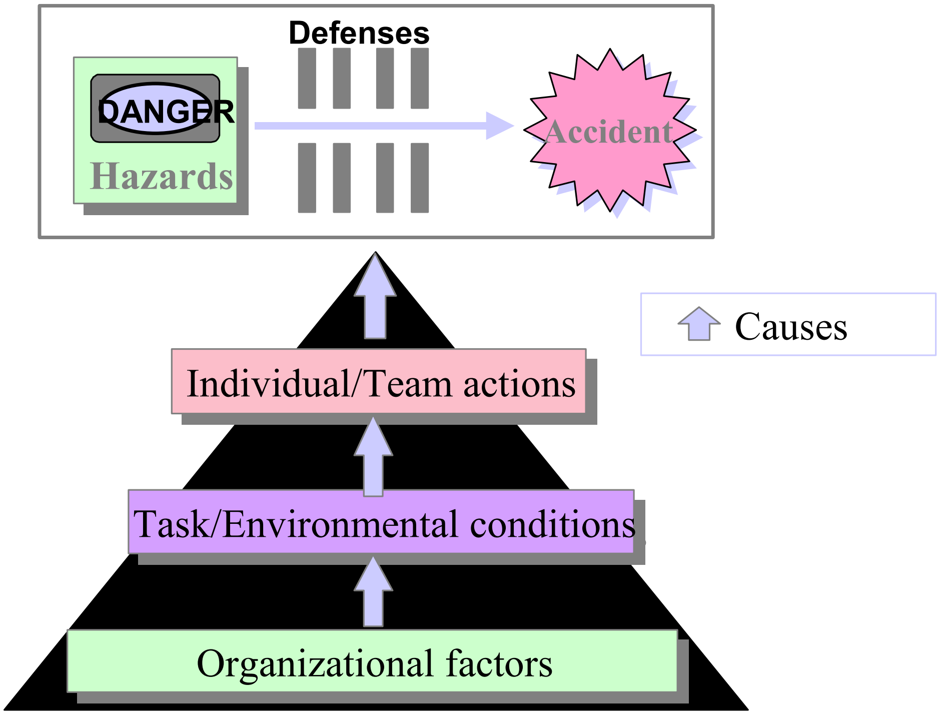
Failures to take action to prevent hazardous events, and actions that cause hazardous events, are commonly called "human errors" in quantitative risk assessments. This term does not imply that people are necessarily personally responsible or culpable in some way, just that an action was omitted (or taken) that adversely influenced safety.

In the context of HRA a human error is simply an action taken (or omitted) by a person that leads to an unwanted outcome—it makes the situation less safe.[1] Note that there is no attribution of blame or fault embedded in this view. People can be placed in situations where an error is almost inevitable. Often we can say "It was not his fault" in regard to some error where we can see almost anyone could make the same error in the same situation. Increasingly, the term "unsafe action" (or "unsafe act"), rather than "human error," is being used in HRA, to emphasize that it is the action (or the failure to act) that is of concern, not whether the action would be considered an error. For example, if a person were led into taking an unsafe action by their training and procedures, many people would say that that was not an error in the normal sense of the term, yet the action had unsafe consequences.

Figure 1 shows a top-level representation of human performance, how human errors can create weaknesses in safety defenses, and how those human errors are conditioned by the environment in which people work. At the very top level, potentially hazardous situations (such as train collisions with other trains and roadway workers and derailments due to overspeeding) are prevented from becoming accidents through defenses being in place. The defenses include the train crew complying with the rulebook of operations, the use of the computer-aided dispatch system (CADS), adhering to speed limits, and the application of fail-safe design principles[2]. For the most part, these defenses prevent accidents. However, these defenses presently rely almost exclusively on human performance—for example, there are very few automated defenses other than the checking effects of CADS in dark territory.

[1] The concept of "human error" has diverse interpretations in the different disciplines of engineering, psychology, and the law. See *Human Error: Cause, Prediction, and Reduction* (Senders & Moray, 1991) for the results of a workshop intended to characterize the different facets of the term.

[2] Fail-safe design seeks to eliminate the hazardous effects of a failure by having the failure result in non-hazardous consequences.



(Adapted from *Managing the Risks of Organizational Accidents,* Reason, 1997)

Figure 1. Relationship of Safety, Human Errors, and Their Influences

Unsafe actions by individuals or teams (such as the train crew) can reduce the effectiveness of the defenses, thereby making the likelihood of an accident higher. It is the purpose of the probabilistic risk assessment (PRA) to estimate the frequencies of such accidents by estimating the probabilities of failure for each of the different defenses. It is the purpose of the HRA task to estimate the probabilities of the human errors that can potentially fail the defenses. However, this estimation needs to take into account the work environment and task conditions under which the work is done, since these can provide an important influence on the likelihood of error. For example, bad weather, long shift times, and high workload all can increase the likelihood of human errors. In turn, work environment and task conditions are often influenced by organizational factors like work rules, duty times, and so on. Therefore, the error estimation process needs account for these *contributing factors*, either explicitly (by the modeling process making adjustments) or implicitly (through using data that already incorporate the practical influence of these factors).

An important aspect of the human reliability analysis process is to identify the *contributing factors* that may cause an unsafe action to be made. Contributing factors can be external (to the person) conditions like poor radio equipment or signals, or a train that is "difficult to control," or internal (to the person) conditions like fatigue or boredom, which we know lead to paying reduced attention to the track ahead. While conceptually separate, in practice these often interact. For example, fatigue is unlikely to cause an unsafe action in a simple routine task, but is very likely to cause an unsafe action in a very challenging situation where concentration or detailed memory recall is required. In practice, we think it makes little difference whether a contributing factor is classified as external or internal. What matters is whether it is a problematic situation.

Therefore, when this report discusses contributing factors, we generally do not concern ourselves with whether they are external or internal.

Further, things can become more complicated when one unsafe action becomes a contributing factor for another. For example, an engineer may mishandle the train, and the resulting behavior of the mishandled train creates the conditions that lead to further errors. In the current example, unsafe actions in train control may lead to an overspeeding train. The engineer, in trying to control the overspeeding train, may make braking "errors" that cause a derail. In other industries, most accidents involve multiple unsafe actions[3].

2.2 Relationship with Risk Assessment Activities

Human reliability analysis is just one component, though a very important one, of an overall PRA such as the type required under the proposed FRA rule. In terms of the relationship between HRA and the PRA, perhaps the most important is that the PRA defines the scope of human errors for HRA required for estimation. The PRA lays out the basic events that can (singly or in combination) result in the hazardous events of concern to the end-user of the study—here, the FRA. "An event," refers to a significant occurrence that has the potential to be an accident in the wrong circumstances. For example, a train being in a block for which it has no authority is "an event." If another train happened to be in the same block traveling in a location and at a speed where it would not see the "intruder" in time to stop, then a collision would occur. The train being in the block may be the result of an unsafe action, such as the engineer failing to recognize the limit of his authority or the dispatcher incorrectly giving the engineer verbal authority to proceed. However, the train could enter the unauthorized block for other reasons, such as mechanical failure of the braking system. Therefore, an event can occur for several or many reasons, some of which are unsafe actions. A *human failure event* refers to an event that occurs as a result (either in part or entirely) of one or more unsafe actions.

The PRA usually specifies what *human failure events* are to be quantified, such as train enters block without authority. Once the PRA has established the overall framework of actions that need to be modeled, the HRA can develop its own internal set of representations of human actions that are consistent with the type of analysis to be performed for the individual human errors. The HRA examines the set of contexts and unsafe actions that can produce that human failure event. In most cases there are multiple different contexts and unsafe actions that can produce the same human failure event. For example, a train can enter a block without authority because of a dispatcher error (e.g., the dispatcher gave the train crew verbal authority to enter the block, but failed to enter the information into the Computer-aided dispatch system). Alternatively the train could enter the block without authority because the train crew was distracted and failed to stop. Yet another alternative is that the train crew intended to stop but underestimated the braking distance required. The function of the HRA analysis and quantification process is to uncover the various contexts and unsafe actions that can result in a

[3] For example, a review of major aviation accidents by the U.S. National Transportation Safety Board (NTSB) showed that in the 37 major accidents reviewed from 1978 to 1990, the range for the number of unsafe actions per accident was from 3 to 19, with a median of 7. [*A Review of Flight Crew-Involved, Major Accidents of U.S. Air Carriers, 1978 Through 1990 - Safety Study* (NTSB/SS-94/01 (PB94-917001)). Washington, DC: U.S. National Transportation Safety Board. 1994].

given human error event, and to quantify the probability of the human failure event, given the variety of contexts and unsafe actions that can lead to it.

2.3 HRA Process

The estimation of the probabilities of human failure events and their contributing unsafe actions can be performed in several different ways. First, various different kinds of models exist to estimate these probabilities and are based on such parameters as the time available for people to take necessary actions, or the quality of indications and instructions for various tasks. Many of these models are summarized by Gertman & Blackman (1994), although more recent developments, such as the ATHEANA method (NRC, 2000) that focuses on cognitive processes and problems, are not included. Almost all of the these developments have taken place in the context of the nuclear power industry and the need to model human actions under extremely rare and challenging conditions, as during a nuclear reactor accident for which few relevant data exist.

A second approach is to recognize that data exist that are related to the kinds of failure events and unsafe actions being modeled. Unlike the actions associated with extremely rare events, these data are usually associated with everyday, or at least frequent, activities like routine train operations, maintenance actions, and so on. Depending on the kinds of data that are gathered, these data sources can be used to identify ranges of probabilities for specific types of unsafe actions.

A third way of estimating human error probabilities is to use the experience of domain experts as a basis for estimation. In particular, the experts need to be experienced in the performance of the tasks being modeled, and the different kinds of errors that can occur under actual working conditions. While, these experts (such as locomotive engineers and dispatchers) may not have expertise to express their opinions in formal statistical terms, techniques have been developed to help elicit their knowledge and convert that knowledge into probabilities, as described below.

These general approaches are not necessarily mutually exclusive. Wreathall, in an evaluation of these different approaches (Wreathall, 2001), recommended that the most useful results for those cases where relevant data exist, is to combine the use of data and expert estimation. The data often provide a broad base for estimation, but almost all databases have limitations and gaps (such as the criteria for events to be recorded) compared with the modeling requirements of the PRA. The expert estimation process provides a way to make adjustments for these limitations and gaps. This overall approach is recommended for studies such as this where an agency like FRA must evaluate base cases and the effects of change.

This section describes the basic steps involved in performing a human reliability analysis. The objective is to describe a general process that can be used to perform an HRA as part of a PRA. The goal is to generate HRA results that are credible, acceptable to the broad set of stakeholders, meet accepted standards for human reliability analysis, and are able to be integrated into probabilistic risk assessments.

The general steps that need to be performed as part of an HRA are:

- Identify the specific unsafe actions to be estimated, as defined by the context of the PRA

- Perform a qualitative human factors analysis to identify the major factors contributing to human risk and reliability.
- Identify the relevant data sources for each action to be modeled
- Identify the limitations and gaps in each data source as related to the actions being modeled
- Implement an expert elicitation process to overcome the limitations and gaps in the data sources
- Synthesize and document the results
- Perform a review of the results by people familiar with train control operations to make sure the analyses and results are compatible with their experience.

Each of these steps is described in more detail below.

2.3.1 Identify the Human Failure Events and Unsafe Actions to be Estimated

In most cases, the PRA will have developed a list of human failure events that it considers as potential contributions to the hazardous events it is modeling. Sometimes these will be identified to the level of unsafe actions. It is recommended that the HRA modeling team and the PRA team jointly review this list and agree to a scope of the HRA modeling that will satisfy the requirements of the PRA.

This list should be developed to identify the particular unsafe actions relevant to the human failure events being analyzed and the level at which they will be modeled. For example, will the modeling separately represent basic “errors” and recoveries or will they be modeled such that only the final outcome state will be represented? Will the model, for example, separately identify the failure of the engineer to recognize the end of their authority and failure of the conductor to correct the engineer’s unsafe action, or will the analysis just model failure of the crew to stop at the appropriate limit? As observed by Wreathall (2001), the recommended practice is to model at the level of the events in the database if possible, since this results in fewer opportunities for mismatches between the data and the modeling. Any lower level or subdivision of modeling should be undertaken only if necessary to generate results that must be used at different places in the PRA. For example, unsafe actions and their recoveries should be separated only if the recovery mechanisms in the situations being analyzed as part of the PRA are substantially different from those represented in the database.

The product of this activity will be an agreed scope of unsafe actions to be modeled in the HRA task, and for which results will be provided to the PRA at the end of the HRA task.

2.3.2 Perform Qualitative Evaluation of Human Factors Issues

Once the scope of failure events and unsafe actions to be modeled in the HRA is defined, the next step is to develop a qualitative understanding of the major factors contributing to human risk and reliability. This involves a human factors analysis of the current work environment, and its impact on human performance.

A qualitative analysis also serves to identify the possible impact of a new technology on human performance and the potential for unsafe actions.

The FRA risk-based evaluation process requires a risk analysis to determine whether the introduction of a new technology, such as positive train control, will result in a level of safety that is equal or higher than the level of safety given current technology. This requires first performing an analysis to quantify the risk associated with the base case (with existing technology) and then comparing this risk to the estimated risk once the new technology is introduced.

A qualitative analysis can identify the major sources of human risk and reliability in the base case. It can also be used to identify the possible impact of the new system on human performance and potential for unsafe actions. The qualitative results can feed into the HRA quantification process and provide additional qualitative information to support evaluation of the proposed new technology (Product Safety Plan).

Evaluating factors influencing human reliability in the current environment

While documents such as operating rules and procedures, and human performance databases, can serve as a starting point for a human factors analysis, these sources often provide an incomplete picture of the actual demands of the work environment and work practice.

A more comprehensive understanding can be obtained through direct observation of the work domain and interviews with the people who are involved in the work (Mumaw, Roth, Vicente, & Burns, 2000). In the context of railroad operations, this means conducting visits to the work sites in question (e.g., dispatch centers, rail yards) to observe the work context directly, and interviewing the people who have direct experience with the job (e.g., locomotive engineers, dispatchers, roadway workers). Useful sources of information include: the workers themselves, labor representatives, first-line supervisors and managers, and training staff.

Observation and interview methods may draw on a variety of methods that include ethnographic approaches (Gamst, 1990; Heath & Luff, 2000; Jordan & Henderson, 1995; Nardi, 1997), cognitive field studies (Roth and Patterson, in press), one-on-one structured interview techniques, or focus group techniques that elicit information from multiple people at once (Krueger & Casey, 2000).

Observations and interviews enable human reliability analysts to uncover and document physical, cognitive, and collaborative demands imposed by the work domain and the strategies that workers have developed to cope with those demands. In many cases these factors and the strategies that domain practitioners have developed to cope with them are not documented or well understood and can only be uncovered by observing and interviewing the individuals directly engaged in the work.

Observations and interviews provide an important source of information about the nature of the work, the factors in the environment that add complexity and create opportunities for error, and the kinds of errors that can occur. This includes an understanding of the broad range of worker duties and practices, the characteristics of the physical environment that can contribute to error (e.g., lighting, temperature, noise), the characteristics of the tools and systems that people interact with that can contribute to error (e.g., characteristics of computer systems, radios), the mental and physical demands of the work itself (e.g., the cognitive demands, the distractions that can arise, the need to time-share tasks), the need for communication and coordination with others within and outside the immediate work environment, as well as the characteristics of the organizational environment (e.g., attitudes, policies, procedures) that can influence performance and contribute to error.

Observations and interviews provide an opportunity to learn about the kinds of errors that have occurred, and the factors that contributed to those errors. It allows the analyst to learn about ‘near misses’ that were never documented since they didn’t lead to a reportable accident.

People not only contribute to increased risk through errors, they also contribute to increased reliability by catching and correcting problems before they lead to an accident. An important objective of the qualitative analysis is to identify the individual, team, organizational, and system factors that enable problems to be caught and corrected before they lead to serious negative consequences.

Recent research across a variety of domains (e.g., aviation, medicine) have shown that highly trained professionals make errors with relatively high frequencies (Amalberti & Wioland, 1997; Reason, 1998). For example, Amalberti reports error rates of up to 3 per hour for aviation cockpit crews are not unusual. The mark of a high reliability system is not that errors are rarely made, but that there are mechanisms in place that enable error detection and recovery. The best pilots and surgeons anticipate the likelihood of errors and develop effective compensatory and error recovery strategies. Similarly, the mark of a high-reliability team is that they are able to catch and recover from each other’s errors. For example, in the studies of the cockpit crews cited by Amalberti, the overwhelming majority of the errors are detected and recovered by the crews in less than 10 seconds.

One of the important aims of a qualitative analysis is to understand the factors in the current environment that enable errors to be caught and recovered. Understanding the factors that make the current system robust to errors in evaluating the potential impact of proposed changes. Changes in technology can have unintended negative consequences. For example, they may eliminate a feature of the current environment that on the surface appears to be of no consequence, but in fact supports robust performance and reduces the potential for error.

Evaluating the potential safety consequences of new technology

One of the proposed uses of human reliability analysis is to support the risk-based evaluation of new technology. When a new technology is introduced that requires human interaction, you cannot evaluate the performance of the new technology in isolation. You need to consider the role that the human may play in either enhancing the overall performance or degrading it. This requires performing analyses to quantify risk in the base case (with existing technology) and comparing this risk to the estimated risk once the new technology is introduced.

In estimating the risk associated with the new technology, it is necessary to consider the impact of the new technology on human performance (Parasuraman & Riley, 1997; Parasuraman, Sheridan & Wickens, 2000). There are a number of human factors issues that need to be considered when evaluating the likely impact of a new system on human performance and potential for error.

One of the first questions to ask is what is the joint ‘person-machine’ system design? This includes:

- What functions will human and machine agents perform?

- What information will be passed among them?
- How good is the performance of the human and machine elements of the system expected to be (e.g., what is the expected accuracy; what is the expected reliability)?
- Whether, and on what basis, one element of the system will be allowed to over-ride or take over from the other?

Other questions to consider include:

- Does the new support system change how the human performs?
- Does the new support system prevent and/or catch and help recover from the types of unsafe actions known to occur in the base system?
- Does the new system introduce any new sources of risk?
    - Does it contribute to any new types of unsafe action?
    - Does it place the human in situations that might encourage them to circumvent it?
    - Does it introduce any other new sources of risk?
- Are there mechanisms built into the new support system that allow the human to play a supervisory control role that would mitigate the potential for any new sources of risk created by the introduction of the system (i.e., opportunities for humans to provide coverage for any new sources of risk)?

If designed well, the joint human-machine system can perform better than either human or machine on their own. If designed poorly, the joint human-machine system can actually perform worse than each of the individual elements. For example, if an automated system has a relatively high miss rate or a relatively high false alarm rate (e.g., in ambiguous or conflict situations), then the human may choose not to use it, or over-ride its decision even under conditions where the automated system is correct (Parasuraman & Riley, 1997). Conversely, the human may accept the recommendation of the automated system even in cases where the automated system is beyond its bounds of competence. The goal is to assign roles to the human and computer elements of a system, and provide them with the necessary information and displays to support these roles, so as to maximize the joint human-machine system performance.

Problems associated with poor joint human machine system design have included:

- Loss of operator vigilance and situation awareness resulting in complacency and an increase in vigilance-associated human errors. As operator confidence in the automatic system increases, the operators tend to become more complacent and less vigilant. Thus, they may fail to detect indications of impending or existing automation problems which require human intervention (Sheridan, Gamst, & Harvey, 1999).
- New opportunities for unsafe actions related to configuring the automation (e.g., inputting wrong values into the automated system such as a wrong ID or destination code)

- Skill loss. With increased supervisory train control technology, the opportunity for the operators to perform the task themselves decreases. The lack of opportunity for practice contributes to skill loss. Skill loss is a problem where the automated system becomes inoperable or is beyond its bounds of competence and the human must take over.
- An increase in workload demands during high tempo high-risk conditions where workload is already very high. One of the common pitfalls of automated systems is that they automate the "easy elements" of a task, reducing workload during periods where workload is already low, but require extensive human intervention for the difficult cases (e.g., aircraft landings) where workload is high.

There are a number of qualitative methods that can be used to evaluate the potential impact of a new technology on human performance. Approaches include:

- A review of the relevant research base both within the railroad industry and in related industries (e.g., aviation, process control). Examples include a review of experiences within the railroad industry with respect to the introduction of new train control technologies such as the Automatic Train Control Systems that was evaluated as part of the Swedish TRAIN-project (Kecklund and the project group, 2001), as well as review of experiences with new automation in the aviation industry (e.g., Woods, Sarter, and Billings, 1997).
- A human factors evaluation of the proposed design or of an early prototype implementation of the design can be performed to assess how well the proposed design adheres to established human factors design principles (Billings, 1997). This can be performed by a human factors specialist with knowledge of problems associated with poor 'joint person-machine' designs and human-centered design principles for effective 'joint person machine' systems (e.g., Christoffersen and Woods, in press; Roth, Malin, and Schreckenghost, 1997).
- Interviews of domain practitioners who have had an opportunity to review and/or use early prototypes of the proposed system. Domain practitioners have operational knowledge and experience that allow them to recognize factors that may limit the usefulness or usability of the system that the designers may not be aware of. Examples include complex cases that the system will not be able to handle, environmental issues such as lighting or noise level that may make the user interface difficult to use, or high workload or multiple attention demands that may make it difficult to use the system as envisioned by the designers (e.g., a locomotive engineer may need to focus his or her visual attention out the window and may be unable to continuously monitor a display for messages and warnings.)
- More formal 'person-in-the-loop' evaluations of the system. These person-in-the-loop tests involve evaluation of the joint 'person-machine' system. The tests examine the ability of domain practitioners to utilize the system effectively in a range of realistic conditions. For example, if there is a new automated system about to be implemented in a locomotive cab, then a person-in-the-loop test would involve having locomotive engineers run a train equipped with the system. Objective measures (e.g., time to detect a system message, time to take necessary action) can then be obtained to assess

the impact of the new technology on performance. The tests could be done either in a high fidelity simulator or in the field.

Generally the qualitative analysis incorporates several of these methods in order to gain a fuller understanding of the issues involved in the introduction of the new technology and the potential impact on human performance.

Summary

Qualitative analyses enable human reliability analysts to more realistically model the types of unsafe actions that occur and the factors that contribute to those errors. Specifically, qualitative analyses enable human reliability analysts to:

- Identify the major sources of human risk and reliability in the base case:
  1. What are the most likely forms of unsafe action in the base case?
  2. What are the factors that are most likely to contribute to those errors?
  3. What recovery mechanisms do humans provide that contributes to a robust, high-reliability system?
- Identify the likely impact of the new system on human performance:
  1. Does the new system prevent and/or catch and recover from the types of unsafe actions that are known to occur in the base system?
  2. Does the new system change how the human performs?
  3. Does it contribute to any new types of unsafe action (e.g., foster complacency, create a source of distraction)?
  4. Does the new system introduce any new sources of risk? Does the system design allow the human catch and recover from the 'system errors?'

Results feed into the HRA quantification process and provide additional information to support evaluation of the proposed system (Product Safety Plan)

2.3.3 Identify Sources of Relevant Data

The process of quantification begins with an evaluation of the relevance of available data to the human actions under analysis. For each of the human actions identified in the list created jointly with the PRA task, it is necessary to identify potentially relevant data sources that can be used to estimate the frequencies with which these errors may occur, and what the number of opportunities may be for such events. Dividing the numbers of errors by the corresponding numbers of opportunities will yield the needed probability of error per occurrence.

It is unlikely that one data source will provide all the needed information. Further, if possible it is helpful to obtain multiple data sources so that several estimates can be created for cross-comparison and selection of a suitable probability range can be made for each unsafe action to be analyzed.

Examples of potentially useful data include the following. Specific additional sources may exist, depending on the particular unsafe actions being analyzed.

- FRA incident databases

- FRA operating experience databases
- Railroad incident databases
- Railroad disciplinary actions databases
- Railroad operating experience databases

The data in the FRA databases are generally available for specific railroads. The railroad-specific data may be available for specific sections of track or territories, or only for the whole system. Issues associated with using these data are discussed next.

2.3.4 Identify Limitations and Gaps in Data Sources

It will be almost certain that the databases identified in the previous step will not match exactly the unsafe actions and events being analyzed in the HRA. Typically, there are two kinds of gaps

1. The database includes events that are not relevant to the kinds of unsafe actions being analyzed
2. The database does not include all events of the type being analyzed in the HRA

An example of the first gap would be the reports of all incidents within a railroad system when the analysis is only concerned with (for example) overspeeding or authority exceedance within one particular type of train control system. In this case, the database must be *filtered* to identify only the events that match the scope of the analysis. Other examples of events requiring filtering of the database include:

- Events associated with dispatchers, roadway workers, or other errors when the analysis is only concerned with train-crew errors
- Events associated with signal territory when the analysis is only concerned with DTC-related events, or vice versa
- Events associated with passenger train control operations when the events of concern can only occur in freight operations, or vice versa.

An example of the second gap would occur when there are criteria that must be met before events are recorded in the database, such as an amount of economic loss or whether there were injuries. Events not meeting these reporting criteria would be missing from the database, even though they are relevant to the HRA study. In these cases, the data from the database must be *scaled* to adjust for the missing data. Other examples include:

- Events associated with a disciplinary database for which there is a significant likelihood that no one would observe the event (and self-reporting is unlikely) resulting in under-reporting.
- Events associated with a disciplinary database for which the error is technically a breach of the rules found during testing but has a negligible impact on safety (such as a few-foot incursion into an unauthorized block)[4]

[4] Note that the decision as to whether such events should be filtered out needs to be made taking account of the PRA and HRA models about what is a meaningful error. It is possible that the PRA is including all rule violations and assessing what fraction is significant within the PRA itself.

Each set of human actions and related databases must be reviewed on a case-by-case basis to determine the specific filtering and scaling requirements. These reviews need to take place in conjunction with people who understand the precise scopes of the databases, and with people knowledgeable about the real-world operations to identify the types of events that may be missing.

2.3.5 Elicit Expert Opinion to Adjust for the Gaps and Limitations in the Databases

The previous step identified the need to filter and scale the data in the databases to adjust for gaps and limitations. In other cases, no relevant data may be available. Both needs are best met by formally eliciting expert opinion. In this case, the experts are the people involved in (or have very detailed personal knowledge of) day-to-day operations that are the focus of the HRA and PRA studies. Relevant experts would include 'front-line' workers (engineers, conductors, dispatchers, roadway workers, etc.) with some operational experience, together with people knowledgeable about the scope and content of the databases. Where the scaling involves making judgments about the relevant operations across an entire company (for example, if operations in the territory under analysis is being compared with the system as a whole) then there need to be experts who are able to make such comparisons based on their experience. Railroad operations management and national union representatives would typically provide such expertise.

There is an extensive history of research on group decision-making. Early work began with development of the Delphi Method.[5] More recent specialization of the elicitation process can be found in many fields. In a major risk analysis effort, updating the risk analysis approach used in the nuclear power industry, the U.S. Nuclear Regulatory Commission sponsored an extensive analysis of a number of reactor plants. The project is known by its main summary report, NUREG-1150 (NRC, 1990). Part of that project included a major effort in expert elicitation. A description of the approach was originally issued as another NUREG report and was later published commercially, (Meyer & Booker, 1991). This book cites a wide range of relevant psychology and operations research literature and includes an overview of that literature with direct guidance and warnings about the pitfalls.

Kahneman, Slovic, & Tversky (1982) describe many of the biases affecting humans in the assessment of probability, such as representativeness, availability, and overconfidence. They also discuss risk perception and procedures to correct for problems in assessment.

More recently, the Nuclear Regulatory Commission sponsored an effort to develop a structured process for expert elicitation to address key uncertainties in the vulnerability of reactors to seismic events. The Senior Seismic Hazard Analysis Committee (SSHAC) produced the most complete, integrated description of the expert elicitation process. Although their domain for elicitation applications was seismic hazards, the elicitation *process* they describe is domain-free and is directly applicable to any elicitation problem.

Their report, known as the "SSHAC" report (Budnitz, Apostolakis, Boore, Cluff, Coppersmith, Cornell & Morris, 1997), offers an effective structure to make the elicitation process consistent. They describe four levels of analysis, from a very simple process to a large group process that is

[5] An excellent source for understanding the many variations of Delphi is the book by Linstone and Turoff (1975). There are chapters on the philosophy of Delphi, numerous applications, evaluations, and potential pitfalls, as well as a wealth of citations covering the history of the technique.

very carefully controlled. There are two important requirements in the process. First the process requires a facilitator to ensure that all participants are heard, all opinions are supported by evidence, and that there is protection against possible unintended bias. Second, the group agrees to seek a consensus position, one that would be representative of the associated technical community. This process implies the inclusion of uncertainty in any estimates. In Appendix J of the SSHAC report, the authors provide a useful comparison of mathematical and behavioral schemes for aggregation of information from multiple experts.

In many cases, the elicitation process is best performed at a workshop where all the experts can be brought together to combine the different sources of knowledge and to make joint estimates for the filtering and scaling. Key aspects of the SSHAC report expert elicitation structure were adopted for the Quantification Workshop conducted for the CBTM case described below.

When all parties fully share the available information (share their evidence), and, when uncertainty is explicitly addressed, consensus can be reached (Bley, Kaplan, & Johnson, 1992). When all parties are forced to explain the basis for their judgments, participants can debate their merits and a consensus distribution can be developed to represent the state-of-knowledge of the analysis team.

A number of controls are incorporated for the following reasons:

- Avoid unintentional bias;
- Force a deliberate consideration of uncertainty;
- Test the reasonableness of distributions developed by the group;
- Search for dependence effects;
- Protect against over-confidence.

These controls address issues of bias raised in the human decision-making and bias literature (Hogarth, 1975; Kahneman et al., 1982; Winkler & Murphy, 1968). Many are described in the SSHAC report. For example, the facilitator must understand how the issues of bias mentioned above affect human assessment of probability and be alert for their symptoms. One good tool for checking the reliability of an assessed distribution is to ask the experts which of two ranges of values of a parameter are more likely. If the facilitator has chosen ranges with equal probability from the distribution and the experts favor one over the other, it is a clue that the group must revisit the assessment. By questioning the group and forcing them to think about unusual conditions (weather, fatigue, time of day, etc.), the facilitator can see if the assessed uncertainty range is broad enough. “Salting” the questions with examples from his or her own experience, the facilitator can encourage the group to expand their thinking.

There are alternative approaches that rely more on testing and rating experts with calculated adjustments of their estimates (Cooke, 1991). These approaches appear especially useful for real-time elicitation (Aspinall & Cooke, 1998), when there is little time to bring all the evidence under scrutiny. The developers of this method suggest using this approach when there is insufficient time or budget to follow a more interactive process. We believe that a group consensus process where experts have the opportunity to examine and discuss the available evidence is more appropriate for developing the human reliability information.

2.3.6 Synthesize and Document Results

The final results are generated by combining expert elicitation results for parameters not represented in available data, expert judgment concerning the appropriate censoring and extension of existing data, and calculations relating large databases to the restricted conditions that exist in the territory of this study.

In all cases, uncertainty distributions are developed to place the results in the context of the full range of issues affecting the assessment team's uncertainty. This distribution includes both randomness and uncertainty related to state-of-knowledge. The resultant probability distributions are generally histograms (rather than analytic distributions), because they are generated through combinations of distributions for various parameters affecting the calculations.

Finally, the tasks described in Sections 3.3.1 through 3.3.5 are documented. This documentation serves the following goals:

- Preserves the list of specific human failure events and unsafe actions that were estimated;
- Shows the context required by the PRA;
- Describes the various data sources that were used for each error that was modeled;
- Shows limitations and gaps in those data sources and how they were handled;
- Provides a record of the elicitation sessions and the associated calculations.

The results are assembled in a concise form for delivery to the PRA team. If possible, the report should provide both complete probability distributions and approximate analytic probability distribution parameters to support the PRA task.

2.3.7 Review

The results of the human reliability analyses, like any other component of a PRA, should be reviewed by members of the team performing the PRA and by a group of people familiar with the train control operations being modeled, to ensure that the scope of the study has been reasonably accomplished and that the results appear reasonable to someone not directly involved in their generation. If possible, the group that performs these analyses should include all parties interested in using the results, such as FRA, the relevant labor groups, and the railroad companies. In the context of the FRA's planned uses of the PRA, a review by the relevant members of the Railroad Safety Advisory Committee (RSAC), such as those who attend its PTC subcommittee, could be appropriate. Such reviews are not expected to evaluate the details of the HRA estimation process but to judge the relative magnitude of the quantitative human reliability results and to assess them against the reviewers' " domain knowledge."

3. EXAMPLE ANALYSIS FOR CBTM STUDY

A major objective of this project was to provide a demonstration of the human factors quantification process as input to risk quantification models such as the Axiomatic Safety-Critical Assessment Process (ASCAP). The CSXT Communications Based Train Management (CBTM) safety case was used to illustrate the methodology.

Using ASCAP, a simulation model was developed to evaluate the potential impact of CBTM on safety. As a first step in estimating the potential impact of a new train control technology, it is necessary to estimate the level of safety in current operations (the base case). The decision-maker can then compare the current level of safety to the level of safety that would be achieved if the new technology were implemented.

Since people play an important role in maintaining safety in railroad operations, it is important to understand how human factors and human reliability influences the overall safety of railroad operations.

The human factors quantification process was used to estimate human reliability values for input to ASCAP. This involved:

- Estimating human reliability values for the base case: direct train control (DTC) operations in the territory where CBTM was tested. This was the CSXT territory between Spartanburg, South Carolina and Augusta, Georgia.
- Examining the potential impact of CBTM on human performance and human reliability when added to the current DTC operations in the above territory.

3.1 What is Communication Based Train Management (CBTM)?

In DTC territory, authority for train movements and track occupancy is accomplished by verbal exchanges between the dispatcher and train crew over the radio. Operating rules govern these exchanges of information between the dispatcher and the train crew. Current DTC operations were used for this ‘base case’ analysis (as defined in FRA’s proposed Standard).

CBTM is a form of train control that provides a warning to the locomotive crew when the train is predicted to exceed the limits of its authority and stops the train if the operator fails to act in time. The system provides four kinds of protection: authority protection, speed protection, work zone protection and switch state protection. The system is intended to provide an overlay safety addition for operations in ‘dark territory’ where DTC is the method of train control operation.[6] When errors occur, such as a communication failure between the locomotive crew and the dispatcher for example, CBTM provides an additional layer of defense. For example, CBTM receives information regarding the authorized train movements from the computer-assisted dispatch (CAD) system used by the dispatcher to indicate valid track occupancy and compares this information with the current train position (using a global positioning system) to determine whether the train is operating within its authority. This system is overlaid over the existing train control system. The train operates under its normal DTC rules of manual operation, with the crew following all the current rules and practices. The CBTM system is intended simply to enforce the DTC rules by applying penalty braking when the train exceeds its block authority,

[6] An overlay PTC system supplements or overlays an existing system of train control.

over-speeds, enters a work zone without clearance, or approaches a monitored switch that is incorrectly set.

3.2 Human Failure Events to be Estimated

The requirements of this study were to analyze the probabilities of specific unsafe actions representing potential contributors to the risks being modeled in the ASCAP study of the CBTM system.

Based on inputs received from the developers of the ASCAP model, the following four events associated with unsafe actions were identified for analysis in the HRA task:

1. Train enters a block without authorization
2. Train exceeds the track speed limit
3. Train enters a preplanned work zone (published in the train bulletin) without authorization
4. Train crosses a misaligned switch.

These events represent the conditions for which the CBTM system can potentially reduce the likelihood of occurrence and fall within the set of functions PTC was intended to address. Therefore, the analysis was performed both for the base case (current operations without CBTM) and the case when CBTM is operational. Other accident scenarios, such as those involving grade crossings or collisions with vehicles used by track inspectors were not modeled because they are not affected by the planned use of CBTM.

3.3 Qualitative Human Factors Analysis

The first step was to perform a qualitative human factors analysis. This involved two aspects:

(1) An analysis of the current work environment to understand the types of errors that can arise and the factors that contribute to those errors;

(2) An examination of the proposed CBTM system, its user interface, and proposed human-system interaction, to assess its potential impact on human performance and human reliability.

An early prototype of the CBTM system was being tested on the CSXT territory between Spartanburg, South Carolina and Augusta, Georgia. This provided an opportunity to directly examine its user interface features and observe its operation, and get input from CSXT locomotive engineers and trainers who had familiarity with the prototype CBTM system.

Two site visits were conducted in support of the qualitative analysis:

1. A site visit was made to the yard in Spartanburg, South Carolina to conduct interviews and observations of CSXT locomotive engineers and conductors, as well as to ride a locomotive equipped with the CBTM system (April 18 and 19, 2001)
2. A site visit was made to the CSXT Dispatch Center in Jacksonville, Florida, to interview and observe dispatchers to understand CSXT dispatch operations and the factors that could contribute to dispatch errors (June 4 – 6, 2001)

The focus of the interviews and observations addressed the following questions:

- What are the most likely forms of unsafe actions in the current railroad operations in the CSXT territory between Spartanburg, South Carolina and Augusta, Georgia (i.e., the base case)?
- What are the factors that are most likely to contribute to those actions?
- What recovery mechanisms do humans provide that contributes to a robust, high-reliability system?
- What impact would CBTM likely have on human reliability and overall safety?
- Could CBTM prevent and/or catch and recover from the types of unsafe actions that are known to occur in the base case?
- Would CBTM change how the people in the system perform (i.e., locomotive engineers, dispatchers)?
- Could CBTM introduce any new sources of risk? If so, are there mechanisms available to enable the people in the system (e.g., the locomotive engineer, dispatcher) to catch and recover from the CBTM 'errors'?

In addition to these generic questions, our observations and interviews were guided by the ASCAP modeling assumptions and human reliability input requirements for the CBTM case, and the specific issues and concerns that members of the RSAC positive train control working group raised with respect to the potential impact of the introduction of CBTM on human performance.

Particular issues raised by the RSAC positive train control working group were:

1. *Potential for complacency*: There was concern that locomotive engineers might grow to over-rely on the CBTM system, and therefore become complacent (Sheridan et al., 1999). In this case, they might become less vigilant in monitoring for conditions where braking is required (e.g., end of authority, speed zones), relying on the CBTM system to provide a backup, should they fail to take timely action. The concern is that if the CBTM system ever fails (without providing any indication that it was not operating), the locomotive engineer, believing CBTM was still operating, might fail to brake in time.
2. *Potential for Intentionally defeating CBTM system*. Another concern that was raised was that locomotive engineers might actively seek to defeat the CBTM system. A concern was raised that locomotive engineers might enter incorrect train consist information in an attempt to change the CBTM braking profile, so that the CBTM system would not activate as designed.
3. *Potential for distraction*. There was also a concern that installing the CBTM system in the locomotive cab would serve as a source of distraction to the train crew. Locomotive engineers would now have an additional demand on their visual attention (the CBTM display), which might serve as a source of distraction, reducing their ability to detect and react to changes outside the cab. These issues were addressed during the interviews and observations.

The results of the interviews and observations provided the background necessary for structuring the topics to be covered in the Human Factors Quantification Workshop, as well as for understanding and integrating the inputs provided by the workshop participants.

The results of the observation and interviews with respect to major error forms and their contributors in the base case, and potential impact of CBTM on human performance were consistent with the inputs provided by the participants in the Human Factors Quantification Workshop.

Section 3.3.1 provides a summary of the interviews and observations of the CSXT locomotive engineers and conductors

Section 3.3.2 provides a summary of the interviews and observations of the CSXT dispatchers

More complete documentation of the results of these interviews and observations are provided in Appendices A and B.

3.3.1 Interviews and observations of CSXT Locomotive Engineers and Conductors

Interviews and observations of CSXT locomotive engineers and conductors were conducted in Spartanburg, South Carolina, on April 18 and 19, 2000. The objective was to form a deeper understanding of the complexities that affect locomotive engineer performance, potential for error, and how CBTM is likely to affect locomotive engineer performance and impact safety.

Activities included:

- Observations during a 4-hour head-end ride on April 19 that was conducted as part of a scheduled CBTM test;
- Two-hour interview with a locomotive engineer that had served as a CBTM trainer, introducing locomotive engineers to the CBTM system.
- Two two-hour focus groups of locomotive engineers and conductors. Eight individuals (six locomotive engineers and two conductors) participated in the focus groups.

The participants in the focus groups were solicited by contacting local labor representatives for the locomotive engineers (Brotherhood of Locomotive Engineers [BLE]) and conductors (United Transportation Union [UTU]) who were informed of the study and asked to put up a flyer that the HRA team prepared announcing the focus group.

The engineers and conductors participating in the focus groups ranged in experience from 11 months to 28 years. They also ranged in experience with CBTM from operating trains with (an early prototype of) CBTM installed on several occasions spanning the period it has been piloted, to having been on only one train run with CBTM installed in the cab. The focus groups were conducted in an off-site conference room and the locomotive engineers and conductors participated voluntarily on their own time.

The interviews/focus group sessions addressed two main topics:

- Factors that make running a train challenging in today's environment and potential for error.
- Potential impact of CBTM on train crew performance.

Factors Contributing to Error

The locomotive engineers and conductors mentioned a number of factors that contribute to exceeding speed limits or going past their limits of authority. Major factors mentioned were attention lapses, distractions, and memory lapses (forgetting).

Train crews were particularly concerned with the possibility of missing temporary speed restrictions (slow orders) and work orders. In the case of temporary speed restrictions, they felt that they were most vulnerable to miss them in cases where:

- The dispatcher issued the speed restriction verbally over the radio after the publication of the train bulletin.
- The speed restriction signs were not put up yet [or were obscured]
- The time duration between when the dispatcher provided the information and when it went into effect was long (e.g., will come to the speed restrictions four hours after the dispatcher called to tell him/her about the speed restriction)

Confusion as to the exact location of a speed restriction or limit of authority was also mentioned as a factor contributing to exceeding speed limits and going past limits of authority. The possibility of confusion was felt to be greatest in cases where the speed restriction or stop location was temporary, especially if in addition: (a) the location was between mileposts; and/or (b) the visibility was poor (e.g., at night or in poor weather) so that visual cues to aid in identification of location was degraded.

Communication errors were also discussed. While communication errors do happen, participants felt that in most cases they are caught and recovered before any negative consequences.

Improper train handling was the last source of error mentioned. The locomotive engineer may know where to stop or reduce speed but overshoot due to braking too late or insufficiently.

Input on CBTM

All eight individuals interviewed in the two focus groups and the trainer felt that CBTM could improve safety. They believed that CBTM could be useful in cases where a train crew might forget to reduce speed or stop at the end of their authority due to attention lapses or memory failures.

They particularly liked that it warns the train crew when they are about to enter a work zone and when temporary speed restrictions are in effect. For these cases, the probability of error is likely to be higher, and the consequences may be severe.

However, all nine individuals also indicated limitations of the current CBTM prototype. Specifically:

- The audio alert was difficult to detect given the noisy cab environment (e.g., engine noise, the whistle, the radio, conversations) and the CBTM visual display was outside of the primary field of view. Difficulty detecting the warning message from the audio alert or the visual display had two consequences:
    1. Failure to respond to a warning message from CBTM resulted in a penalty brake application.

2. Because of the severe consequence of missing an information message or warning (i.e., the penalty brake application,) the locomotive engineers felt a need to continuously monitor the CBTM display. This behavior added significantly to their workload inside the cab. This behavior could potentially distract them from attending to events outside the cab (e.g., trespassers, motor vehicles at grade crossings).

- In many cases, the warning message did not come on early enough before the penalty brake is applied to allow the locomotive engineer to respond in time to avoid the penalty brake. The short time between the warning and a penalty brake application reduced the ability of the locomotive engineer to take advantage of the warning message. It also reduced the ability of the train crew to catch and recover from any ‘errors’ that the CBTM system might make. Thus, it reduced their potential to serve as a recovery mechanism. More time is needed between the on-set of the warning and the initiation of the penalty brake to allow the locomotive engineer time to slow down the train to the appropriate speed and/or select an appropriate stopping place.[7]
- Often the CBTM system determined that braking was required at an earlier point than the locomotive engineers would choose. In some cases, a warning came on in situations where the locomotive engineers felt stopping was unnecessary or inappropriate. In other cases, the position where the CBTM stopped the train was inconvenient, making it hard to restart the train. Stopping at an inappropriate time or place may also introduce a new source of risk.

With respect to the potential for complacency and over-reliance on CBTM, the engineers provided mixed comments. On the one hand, they indicated that it remained their responsibility to make sure that no movement authorities or speed restrictions were violated, independent of whether they were reminded by CBTM or not. The analogy one locomotive engineer gave was to an advanced warning board on the side of the track. If it is there, it can remind the locomotive engineer of the need to brake soon. However, if for some reason the warning is not there, the engineer is still responsible for braking. The same would be true for the CBTM system. It would provide an aid, but the engineer still bears the responsibility for safe train operation. At the same time, the engineers noted that if the CBTM system were working well they would tend to rely on it. As one engineer put it “If we can’t rely on it, I don’t want it up there. If it works, I’ll rely on it.”

With respect to whether CBTM would change the behavior of the locomotive engineers, the locomotive engineers indicated that it would. Given that the CBTM system expects the locomotive engineers to brake earlier than they are now inclined to, they would need to learn new braking styles. Thus, CBTM raises a need for training not only on the CBTM interface and how to use it, but also training on train handling and braking that is more consistent with the expectations of CBTM.

Locomotive engineers reported that the interface for entering consist information into CBTM was easy to use. When asked whether locomotive engineers might intentionally enter incorrect

[7] There are drawbacks associated with presenting the warning message too early as well as too late. The appropriate length of time required between the onset of the warning and the initiation of the penalty brake can best be determined by conducting empirical tests using locomotive engineers.

consist information in order to manipulate when the CBTM system came on, all 9 individuals interviewed felt that that was very unlikely. They indicated first that CBTM contributed to safety and they wouldn't want to take action to defeat that, and second, since it is a computer system, it records all inputs, and it would therefore be easy to catch when someone did this.

3.3.2 Interviews and observations of dispatchers

A site visit was made to the CSXT Dispatch Center in Jacksonville, Florida on June 4 – 6, 2001. Appendix B summarizes the results of the site visit. One objective of the site visit was to understand current dispatch operations and the kinds of errors that dispatchers were likely to make. A second objective was to obtain feedback from dispatchers and managers of dispatchers on the potential safety benefits and drawbacks of installing CBTM. Both objectives supported the goal of providing human factors input to the ASCAP simulation model comparing the base case to CBTM. The site visit included observation of dispatcher operations as well as interviews with dispatchers, dispatcher training instructors, and managers.

Observations were made at three different dispatch desks that handled primarily dark territory, including the dispatch desk that handles the territory from Spartanburg, South Carolina to Augusta, Georgia, where the prototype CBTM system was tested. In addition, we observed and interviewed a chief dispatcher, whose territory included the territory from Spartanburg to Augusta. In total, we observed and/or interviewed seven railroad dispatchers, one chief dispatcher, two dispatcher-training instructors and two managers of the dispatch center.

Opportunities for Errors and the Contributors to Error

Observations and interviews resulted in a deeper understanding of Dispatch Center operations and the Computer-Aided Dispatch (CAD) system that dispatchers use to enter block authorities. In particular, we were able to identify the most common types of errors made by dispatchers and the factors that contribute to those errors.

Dispatchers provided extensive information on the types of errors that they made and the factors that contributed to those errors. In many cases those errors resulted in a discrepancy between what was entered in the CAD system and what the receiver of the message over the radio believed was the case. For example, the dispatcher can make a data entry error in the CAD system. The dispatcher can verbally say the right thing to the person over the radio but enter the wrong thing in the CAD system. The dispatcher can also verbally give more block authority than he/she enters in the CAD system. A problem can arise if later the same dispatcher or a different dispatcher gives the blocks that were verbally authorized to the first train but not entered in the CAD system to a different train.

Another type of error discussed was communication errors, especially errors due to poor quality radio reception. Examples mentioned include:

- Can mistakenly believe you are talking with a different train;
- Can 'hear' the wrong thing due to noisy radio (static, cut-out) where the listener 'fills in' the missing information based on expectations;
- Can fail to catch errors made by locomotive engineer during readback because the dispatcher moved on to the next task and/or because the dispatcher is also subject to the impact of expectations on perception;

- Locomotive engineer may write something different from what the dispatcher said, but repeat back (correctly) what the dispatcher said.

Most errors are caught and recovered before they have any safety consequences. The individual making the error can detect it (e.g., data entry errors), the person with whom they speak with can detect the error (e.g., a communication error), or a third party can catch it.

Input on CBTM

We were also able to get feedback on the perceived usefulness of CBTM in improving safety from managers of the Dispatch Center, training instructors, and dispatchers. Among the points made were that CBTM:

- Would stop a train if the dispatcher has not put in the block authority information in the CAD system (i.e., in cases where due to 'data entry error' or verbal misunderstanding between the dispatcher and the locomotive engineer, there is a discrepancy between what was said verbally to the locomotive engineer and what was entered in the CAD system);
- Would stop the train if a switch was inadvertently left in the wrong position;
- Would stop a train if it exceeded a speed restriction (e.g., in cases where the dispatcher failed to communicate a temporary speed restriction.)

3.4 Quantitative Analysis

The primary tasks in the quantitative analysis of the HRA were:

- Identification of relevant data sources;
- Identification their limitations and gaps;
- Application of the expert elicitation process to compensate for these limitations and gaps.

3.4.1 Overall Analytical Process

This section describes the analytic process that was used to quantify each of the human failure events specified in Section 3.2. The analytical process for each of the human failure events was intended to answer the following five questions. The first three questions were answered in large part before the quantification process was started, by defining the scope of the analysis and in the discussions undertaken as part of the qualitative analyses.

1. *What are the major unsafe actions to quantify?*
   For example, the train exceeds its limits of authority. This event could be the result of two different general unsafe actions—errors by the train crew and errors by the dispatcher
2. *What is the scope?*
   For each major unsafe action identified in the previous step, what specific unsafe actions are included? In the case of the train crew, the scope would simply be the crew fails to stop the train at its limit of authority. The unsafe action may occur

because the train crew fails to notice when they reach the end of the last block or erroneously recall the limit of their authority.

3. *What factors could cause the actions listed in the previous step?*
   Examples of factors that could cause failure of the train crew to stop at the limit of the authority include:
   - Inattention or failure to recognize their location
   - Erroneous recall of authority limits
   - Distraction (within the cab or outside the cab)
   - Over-reliance on another crewmember
   - Misjudged braking performance
4. *What data exist?*
   To what degree do the available databases relate to the actions being modeled? Do they include all or most of the identified significant causes? Data are needed for both the opportunities for the events and the events themselves.
5. *What judgments are needed?*
   Are there additional causes not included or under-reported in the databases that are relevant to the analysis? For example, are all the causes listed above (inattention, etc.,) included in the train crew disciplinary database or the FRA incident database? Are there additional causes in these databases that should be excluded? On what basis can the data be filtered and scaled?

The final probability parameters (usually in the form of probability distributions) for the unsafe actions are estimated, based on the numbers of events (the numerator) divided by the opportunities for the events (the denominator) The numerator and denominator are each adjusted based on the judgments and databases. This overall process is shown in Figure 2. It shows the routes by which the probabilities of different actions are calculated (at the bottom of the figure) based on data and judgments discussed above.

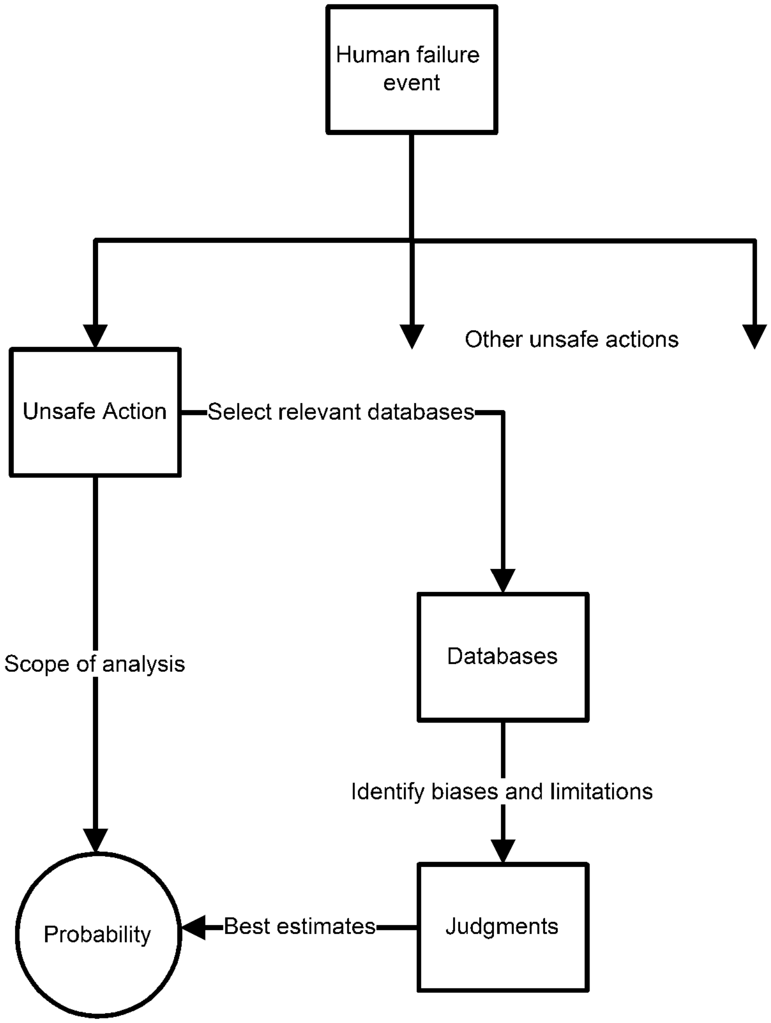


Figure 2. Overall Analytical Process

3.4.2 Sources of Data

Two kinds of data are required in HRA studies: information about the numbers of events similar to those being modeled, and information about the number of opportunities for such events such that a probability of the events can be estimated. Two major sources of data were identified in this study: the databases maintained by the FRA, and databases maintained by CSXT. Both data sources contain information about both the frequencies of events and the opportunities for such events.

Table 1 shows the potential relevance of each of these databases to the events being analyzed. Their limitations are discussed in Section 3.4.3.

FRA Databases

Incident Database

FRA maintains a database that contains coded summaries of incidents in railroads that meet the FRA reporting requirements (see below). These summaries identify the railroad(s) and locomotive identifiers involved, the date and location, the type of traffic (passenger, freight, etc.) and a set of cause codes for the event, based on the investigation made by the railroad. These data can be accessed at the FRA website (http://safetydata.fra.dot.gov/officeofsafety). It is possible to create and download the results of searches of the database. For example, the search may specify the railroad, a date range, and the types of cause codes (say, all those reported as involving human errors) for the events. The results are then available for downloading and analysis offline.

Operational Exposure Database

The operational exposure database maintained by FRA is located at the same web address as the incident reports (above). This database summarizes the amounts of train movements (expressed in train-miles) for each railroad, separated by train-miles in yards versus track, passenger versus freight, etc. Totals per calendar month are provided.

CSXT Databases

Three databases were identified by CSXT staff in discussions about the needs of the HRA study, and were identified on a case-by-case basis. It is possible other CSXT databases exist for other management purposes, but were not identified to this team. As described below, the existing databases were sufficient for these analyses.

Incident Databases

CSXT identified three databases suitable for this analysis. The first was a summary of events that occurred on the test territory between Spartanburg and Augusta in the nine years preceding the study. Each event included a summary of the type of event, and whether it was hardware, human, roadway or other-related. There were 89 events, of which the largest contribution was from roadway problems (e.g., wide gage due to defective or missing ties). A total of 24 events involved a human-related cause. Of these 24 events, the largest number was associated with failure to remove a derail (a total of 4 in the nine years).

The second set of incident databases was associated with disciplinary actions on CSXT. Two were provided: one associated with train crew disciplinary actions and one with dispatcher actions. No individuals were identified in either database. Both databases provided a brief summary of the event (either the rule violated [dispatcher data] or the type of event [train crew data]) and the date. From these data, it was possible to identify if the action was a track segment violation (exceeding the limits of their authority) for a train crew, or if the action was a rule violation concerned the inappropriate issuing of a block authority by a dispatcher.

Operational Exposure Database

For operational exposure, CSXT provided a set of “raw” data for the test territory: a set of all movement authorities for the territory covering a two-week period. This two-week period was considered to typify operations for the test territory. These were provided in electronic format and were converted to a Microsoft Access® database. For this two-week period the following items were identified:

- The number of trains traveling the territory,
- The number of authorities that were issued,
- At what time and the number of blocks issued or released,
- The number and duration of temporary work zone restrictions in place, and
- The number of track inspections occurring at any time.

Table 1. Summary of Data Associated with Each Human Failure Event.

Event	Event data	Unsafe actions	Unsafe action data	Operational exposure
Train fails to stop at boundary of authority	• CSXT incident data • FRA incident data	Train crew fails to stop at block boundary at end of authority - block sign present - block sign missing	• Employee disciplinary actions database	CSXT total & yard miles
		Dispatcher fails to protect train authority in CADS	• Dispatcher disciplinary actions database	CSXT total & yard miles
		Train crew mishears dispatcher as to limit of authority (location, train ID)	• Estimates (e.g., estimates obtained from train crews or dispatchers)	Number of train movements (dispatcher printouts, tables in ASCAP Appendix C)
		Dispatcher mishears train crew request for authority (location, train ID)	• Estimates (e.g., estimates obtained from train crews or dispatchers)	Number of train movements (dispatcher printouts, tables in ASCAP Appendix C)
Train enters work zone without authority	• CSXT incident data • FRA incident data	Train crew fails to stop at work zone boundary - fails to identify location - sign present - sign missing - recognizes location but fails to stop	• Employee disciplinary actions database	CSXT total & yard miles
		Dispatcher fails to identify work zone boundaries (not in train bulletin)	• Dispatcher disciplinary actions database	CSXT total & yard miles
		Train crew mishears or misunderstands Employee In Charge (EIC) communication & proceeds into work zone without authorization	• Employee disciplinary actions database • Estimates by train crews & roadway workers	CSXT total & yard miles

Table 1. Summary of Data Associated with Each Human Failure Event (Continued)

Event	Event data	Unsafe actions	Unsafe action data	Operational exposure
Train exceeds speed restriction	• CSXT incident data • FRA incident data	Train crew exceeds speed restriction - permanent speed zone - sign present - sign missing - temporary speed zone - sign present - sign missing	• Employee disciplinary actions database	CSXT total & yard miles
		Dispatcher fails to identify temporary speed zone (not in train bulletin)	• Dispatcher disciplinary actions database	CSXT total & yard miles
Train runs over wrongly positioned switch	• CSXT incident data • FRA incident data	Previous train crew leaves switch in wrong (“reverse") position without agreement from dispatcher	• Employee disciplinary actions database	CSXT total & yard miles
		Dispatcher fails to warn of switch left in “wrong” position (if agreed with previous crew)	• Dispatcher disciplinary actions database	CSXT total & yard miles
		Train crew does not recognize switch in wrong position and stop – target OK	• Employee disciplinary actions database	
		Train crew does not recognize switch in wrong position and stop – target not OK	• Estimates	
Train crew fails to act on CBTM warning	• None	Crew fails to react in time to prevent unnecessary CBTM action	• Estimates	
Train crew over-relies on CBTM for train control (complacency)	• None	Crew fails to take manual actions to control train in CBTM controlled situations	• Estimates	
Train crew enters wrong information in CBTM for consist	• None	Crew erroneously enters wrong data via CBTM interface	• Estimates	
		Crew purposefully enters wrong information to adjust braking profile	• Estimates	

3.4.3 Limitations and Gaps in the Databases

Each database has certain limitations and gaps with regard to the events being analyzed. The following are the primary instances for which an analysis was performed during the expert elicitation.

FRA Databases

Operational Exposure Database

The FRA operational exposure database provided a basis for estimating total train movements within a given railroad, but the categories only describe whether the movements were in yards or out of yards. There is no distinction between the types of train control system (e.g., DTC) in use, or any information about traffic within specific territories.

Incident Database

There are two primary limitations in this database: events to be recorded must meet certain damage (greater than $6,500.00 in 1997) or injury criteria as set forth by FRA,[8] and the reporting railroad required to provide only a limited set of causal information.

As a result of the first limitation, there is a significant gap of event information for which no accident occurred—there is no “near miss” reporting for events involving errors but no consequence, for example, in the FRA database system.

Because almost any accident is the result of multiple causes, the ways in which an event is reported can be somewhat subjective as to what is given a primary focus: equipment or human. Therefore relying only on the cause codes of the events does not provide a sufficient basis for identifying events relevant to this study. The reports do provide the opportunity for presenting a narrative for additional information but there can be quite significant latitude in the way events are reported. However, the combination of types of events and the narratives seems to provide at least a basic starting point for identifying relevant events.

CSXT Databases

Operational Exposure Database

The details of the traffic were sufficient to identify the total numbers of trains, the numbers of blocks issued, the amount of maintenance work, etc., in the test territory for the two-week period. The only question is the extent to which the two week period was representative of traffic overall in the test location. Summary data for a second period were obtained that indicated a somewhat higher volume of traffic. This second set of data were used for a sensitivity analysis in one of the evaluations to identify what the effect would be if alternative data were available.

Incident Databases

The small number of events limited the CSXT database associated with incidents in the Augusta-Spartanburg territory. The databases associated with disciplinary actions were limited largely by the fact that that an unknown number of similar events could occur, but without any mechanism to detect and report the event outside of the crew involved. In the absence of any incentive to

[8] See FRA Instructions for preparing FRA Form F 6180.54, Rail Equipment Accident/Incident Form

self-report such events, this number cannot be known for certain. On the other hand, some disciplinary events (particularly for train crews) could be the result of performance testing that is more rigorous than normal operations. Therefore, the potential exists for both underestimating (from unseen events) and overestimating (from the inclusion of non-representative testing) from these databases.

3.4.4 Expert Elicitation Process: Quantification workshop

In order to compensate for the limitations and gaps in the available databases, the data need to be filtered and scaled. To perform these adjustments, a two-day elicitation workshop was held on October 29 and 30, 2001 in Greenville, South Carolina, to obtain adjustments to the data represented in the databases available for quantification. Prior to this workshop, a trial workshop was held in August 2001 at FRA offices in Washington, DC, to test the elicitation process and to identify any additional databases that could be used. This trial led to obtaining the set of authority data for the two-week period from CSXT, for example.

A total of 30 attendees participated in the two-day workshop in Greenville. The participants consisted of:

- Four railroad representatives and associated consultants
- Thirteen workers, union representatives and associated consultants
- Six FRA representatives and associated consultants
- One University of Virginia (ASCAP contractor) representative
- Six Volpe Center and associated consultants (including the HRA team).

In order to accommodate this large number, the attendees were divided into two groups for most of the analyses, though the training and the analyses of two particular scenarios were performed with the group as a single body. [9]

In order to ensure that all attendees had a common understanding of the purpose, approach and tasks of the workshop, a training period was provided that covered:

- The goals of the analysis (both the purposes and scopes of the ASCAP modeling and the related HRA activities)
- The approach and tasks being undertaken by the HRA study (using a combination of databases and judgment)
- The technology and planned application of the CBTM system (by CSXT).

Following the training, the group as a whole analyzed one event (train exceeded block authority because of an error by the train crew) as an exercise to see how the process worked in practice. This example is described later to illustrate the analytical process.

[9] The term, scenario, is used in PRA and HRA modeling to describe a combination of equipment conditions and unsafe actions that result in an accident or other situation of concern to safety. For example, a train entering a workzone without authority would be a scenario. It could result from unsafe actions (failure of the crew to obey their limits of authority, or it could be the results of equipment failures (complete brake failure). The scenarios we analyzed were the result of unsafe human actions.

A facilitator who was familiar with the event being analyzed, the types of databases available, and the expertise in the group led the elicitation process for each human failure event. The groups were divided so that expertise for particular events was contained in the group making the evaluation. For example, the group performing the evaluation of dispatcher error events included both dispatchers and locomotive crews. Preliminary analyses were performed to identify talking points in the facilitated groups, such as possible causes of human error events (based on the results of the qualitative analyses described earlier), examples of the databases and their possible limitations, and questions to help develop distributions associated with the probabilities being estimated (such as "How high and how low could the end points of the distributions be, and why?"). The facilitator led the discussion through the items listed in Section 3.4.1, including those that had been prepared (such as the list of causes for the events and the possible databases) to ensure that no significant contributors or sources of information had been overlooked, and that everyone in the group had a common understanding of the events and factors being analyzed. In all cases, an extensive discussion ensued that often clarified the details of the actions necessary for the event to occur, anecdotal examples of near-misses that participants had witnessed or participated in, and issues associated with the content of the databases.

Following this discussion, the facilitator led the discussion to actually estimate the parameters of the model, including uncertainty. Two approaches were needed, one for cases where data were available, but not quite appropriate and a second for cases where no data could be found. In the first type of discussion for estimating what adjustments were necessary to make the database most relevant, the group considered such issues as which events needed to be excluded, and where under-or over-reporting could occur. Based on these discussions, the facilitator led those group members who had a working knowledge of the situation to estimate adjustments to the parameters developed from the database. In almost all cases, these values were obtained by eliciting the endpoints of the distribution (using the "How high…?", and "How low…?" questions) and then estimating the shape and the resulting mean of the distribution. In the second case, where no data applied, the people with hands on experience were pushed to think through how relevant situations could occur, what factors would be most important, and then to focus on the extreme values – what is the most often it could occur and the least often? In some cases the facilitator had to synthesize the discussion, saying, "From what you have discussed, it appears that the high and low values must be…" Sometimes participants with experience on other railroads could suggest things they had seen elsewhere and that would lead to deeper discussion on the possibilities at hand.

This process was followed for almost all of the events being analyzed. The one set where a slightly different process was used was the estimation of the use of the CBTM system. Because this system is still in trial use, and its precise parameters that could alter the way people use it have not been fixed yet, it was necessary to ask the attendees to imagine how it would work in practice, and then the facilitators led the entire group through consideration of certain issues associated with CBTM that were discussed in the qualitative analyses (Section 3.3). Because some of the attendees had used the system in the trials currently under way, this was considered a feasible way to proceed. (More extended applications or use in a simulator could provide additional inputs.) This analysis also provided a setting to discuss current problems in the trial operation (such as the limited warning times before braking) mentioned earlier. Because of the limited experience with the system, and the fact that it was not finalized yet, certain assumptions were made about the final system and its modes of operation. For example, it was assumed that the problems exhibited by the early prototype version of CBTM would be resolved prior to final system implementation. Based on these assumptions, it was possible for the group to provide estimates of the likelihood of particular responses when using the system. These assumptions are discussed in more detail in Appendix C.

3.5 Results of CBTM Analyses

This section presents a worked example to illustrate the analysis and quantification process. The detailed results for the remaining events analyzed are presented in Appendix C. A summary of the results is presented following the worked example.

3.5.1 Worked Example

Train-caused Block Boundary Exceedances

The following example demonstrates the elicitation process and quantification for one event: the event involving a train entering a block for which it does not have authority because of an error by the train crew. The workshop attendees analyzed this event as a single group. Figure 3 shows the overall process used for this analysis.

Scope of analysis

The workshop attendees agreed that the scope encompassed where the train crew fails to stop the train before it enters a block for which they have no authority. Selection of the scope was based on several considerations. Since this study was concerned with the effects of the CBTM system on safety, emphasis was placed on the events where the operation of CBTM can make a difference to the level of safety. The selection of the specific human errors or analysis was made on the basis of whether CBTM would be likely to impact the frequency of such events. The errors included in the following analysis were judged by the analysis team and reviewed by the workshop participants as representing the most likely to be affected by CBTM for exceedance of the block boundary. Equipment faults were not included (such as brake failure). The train erroneously entering a work zone was considered in a separate analysis.

The following error forms were identified by the train crewmembers at the workshop:

- Fails to recognize location due to

  Weather

  Lack of experience

- Misunderstands authority

  Distracted while receiving authority

  Expected to get greater number of blocks than actually issued

  Boundary relocated (occasionally happens)

  Mishears authority

- Distraction or over-reliance by one crew member on the other

  Within the cab (alarms, other tasks, conversation)

Outside the cab (fire, trespassers, etc.)

- Misjudgment of braking

  Normal braking conditions

  Abnormal braking conditions (slippery track)

- ‘Unconscious’

  Highly fatigued

  Environment (e.g., chemical release)

  Drug and alcohol

Relevant Databases

The workshop attendees agreed that the following databases were relevant to this event:

- CSXT train crew disciplinary actions database
- CSXT incident database for the CBTM trial territory
- FRA operating experience database
- CSXT sample authorities database

Analysis

Based upon the available data sources, two paths potentially existed to analyze the likelihood of crew exceedances. One was to use the CSXT disciplinary data that were associated with all CSXT operations, and the second was to focus on the experience within the trial territory. Both paths are shown in Figure 3.

CSXT-wide Analysis

For the crew disciplinary actions, the relevant category is the track segment violations. Ninety-one track segment violations occurred in the 4-year period from 1997 to 2000. The workshop participants agreed that this 4-year period was largely representative of current operations. It was felt that conditions prior to this period were less likely to be representative of current conditions because of updates in the rulebook, company mergers, etc.

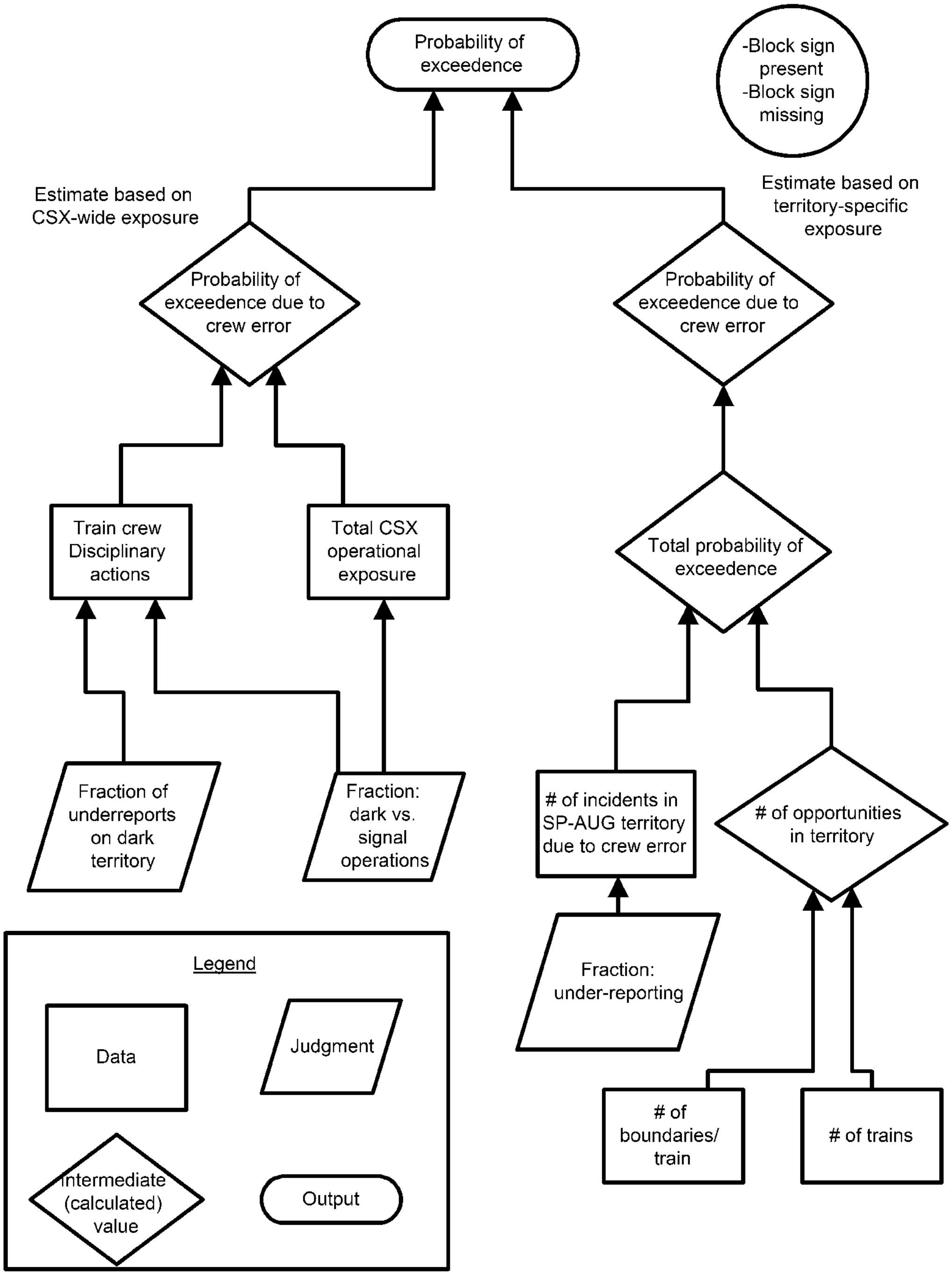

Probability of exceedence
-Block sign present
-Block sign missing
Estimate based on CSX-wide exposure
Estimate based on territory-specific exposure
Probability of exceedence due to crew error
Probability of exceedence due to crew error
Train crew Disciplinary actions
Total CSX operational exposure
Total probability of exceedence
Fraction of underreports on dark territory
Fraction: dark vs. signal operations
# of incidents in SP-AUG territory due to crew error
# of opportunities in territory
Fraction: under-reporting
# of boundaries/ train
# of trains
Legend
Data
Judgment
Intermediate (calculated) value
Output

Figure 3. Analytical Process for Crew Exceedance Event

The 91 track segment violations could have occurred in signal territory, DTC territory, or in yards. No information is provided in the database to distinguish between these locations. Based on discussions within the group, it was agreed that these 91 events could be distributed on the basis of the train mileage associated with each of the territories and yard traffic.

The workshop participants agreed that the basis for these different train miles could be estimated from a combination of two sources of data: the FRA operating experience database (which provides train mileage data associated with track separately from yards), and the relative track lengths associated with signal, DTC, and other operations. Table 2 shows the data from the yard versus mainline traffic over the 4 years analyzed using FRA data.

Table 2. CSXT Train Miles (from FRA database)

Year	Total Train Miles	Yard Train Miles	Non-yard Train Miles
1997	83,733,024	13,324,933	70,408,901
1998	83,447,524	13,367,246	70,080,278
1999	105,277,723	16,075,426	89,202,297
2000	114,426,120	17,874,254	96,585,866

Based on data provided by FRA, Table 3 shows the track lengths associated with different modes of operation.

Table 3. Track Lengths for Different Operating Modes (FRA data)

Territory	Length (miles)	Fraction
Yard*	2,963	13.6%
ATC	75	0.3%
Signal	10,560	48.4%
Dark	6,072	27.8%
Unknown	2,164	9.9%
Total	**21,834**	

* Yard miles were computed to correspond to the relative train miles listed in Table 2.

The "unknown" category represents particular CSXT tracks associated with specific local operations (assumed to be industry-related tracks), and where insufficient information existed from available sources to confirm the specific mode of operations. Because we did not have further data to estimate specifically these tracks as either DTC or signal operations, we modeled them as a source of uncertainty.

The uncertainty distribution was created by setting the two known endpoints to be the fraction of DTC track length out of the total, if (1) none of the unknown track length were DTC (DTC=27.8%), and (2) it was all DTC (DTC + unknown = 27.8 + 9.9, or 37.7%). The most probable value was defined to be if the modes of operation of the "unknown" track length were distributed in the proportions of the known signal and DTC track lengths (30.5%). The ATC and yard operations are a special mode whose locations are known. The result is a triangular distribution for the fraction of CSXT track length that is DTC operations whose mean is 32%. The triangular distribution shape is often chosen when the end points are fixed and known, and there is a plausible logic to identify the peak. It must be noted that the specific shape of the distribution has a very small impact on the result. For example, using a skewed normal distribution with truncated tails at the above end points has a negligible effect on the calculated mean value. This distribution, like the others discussed below, was developed during the workshop, discussed with the workshop participants, and agreed with them as a reasonable basis for the analysis.

The second adjustment assessed whether the 91 track segment events in the disciplinary database represented an under- or over-reporting of the events in practice. The focus of this analysis was a discussion with the group members. It was generally agreed that the events in the database under-represented the events in practice. The experts in the group, the train crewmembers and those familiar with the CSXT disciplinary process, were asked how low and how high the under-estimation could be, using the elicitation process described earlier. The lowest under-reporting was estimated to be 5%, and the under-reporting could be as high as 20%. All intermediate points were judged equally likely, thus creating a flat distribution between the limits of 1.05 and 1.20 for the adjustment to the number of events in the database.

Based on the above analyses, the adjusted number of track segment events/year occurring in DTC territory is:

= (91 x Dist (Fraction of DTC operations) x Dist(Under-reporting adjustment))/4 years

This computation was processed with the distributions described above using Microsoft Excel® and an Excel add-in program called Crystal Ball®.[10] Crystal Ball® computes distributional analyses by simulation, rather than using single numbers, within Excel spreadsheets. This analysis resulted in a distribution that can be represented by a lognormal distribution with a mean of 8.19 track segment events/per year (standard deviation of 0.616) occurring in CSXT DTC territory.

The data in Table 2 and the distribution of the fraction of CSXT total train miles representing DTC operations were used to estimate the average annual number of train miles associated with DTC operations. The total train-mileage averaged over the 4 years was 96,729,600. The resulting distribution had a mean of 26,126,000 DTC train-miles/year. Dividing the 8.19 events/year by 26,126,000 train-miles/year yields a mean of 3.14 x 10^{-7} events per train-mile. The resulting distribution is best characterized by a uniform distribution, with a minimum of 2.47 x 10^{-7} and a maximum of 2.82 x 10^{-7} events per train-mile.

Territory-specific Analysis

The relevant databases for this analysis are the CSXT territory-specific incident data and the sample authority data. The event data for the territory indicates that in the period from 1992 to 2000, no incidents involving authority exceedances were recorded. An extended discussion took place among the group as to the likely range of events that could have occurred in the territory without being recorded. The workshop attendees concluded that over a 10-year period, at least three such events but no more than six events were likely. Values outside this range were plausible. Based on these estimates, a distribution was created to represent the number of events. This distribution was created using a normal distribution having the 5-percentile point of the

[10] Decisioneering, Inc., Denver, Colorado, USA.

distribution set at 3 and the 95-percentile point set at 6, thus allowing for the possibility of values outside the range. Its mean value is thus 4.5/10 years, or 0.45 per year. The selection of the normal distribution (as opposed to the triangular distribution used earlier) was made because there were no hard limits or endpoints (particularly on the high side). In principle, there could have been any (non-negative) number of events. There is no natural upper limit, just a decreasing likelihood of such events occurring. Use of the normal distribution allows for that possibility. Other unbounded distributions could be used, but selecting one with a different shape would make very little difference to the final calculated parameters. As with the selection of other distributions, the selection of this distribution as an assumption was discussed and agreed upon by the workshop participants.

The operating experience was represented by the sample of all territory authorities issued over a 2-week period (6/10-23/2001) for the Spartanburg and McCormick territories (the test area for the CBTM system). In this period, authorities were issued to 273 unique train identifications. The track mileage covered by these authorities is 120.5 miles. Therefore, in the course of the two-week period, there were approximately 3,290 train miles of operating experience. Allowing for 26 two-week periods in a year, this exposure metric corresponds to about 855,310 train-miles per year, or 8,553,100 train miles in 10 years.

The rate of authority exceedance per train mile is therefore calculated by dividing the distribution of Authority Exceedances (/10 years) by 8,553,100 train miles, which is a similar shaped distribution having a mean of 5.26 x 10^{-7} per train mile.

To convert this value to the rate per block authority, the distribution of the number of blocks issued per authority was analyzed. The resulting distribution is shown in the histogram in Figure 4. While the majority of authorities were issued for one block (53%), the mean number of blocks issued was slightly over two, as shown in Figure 4.

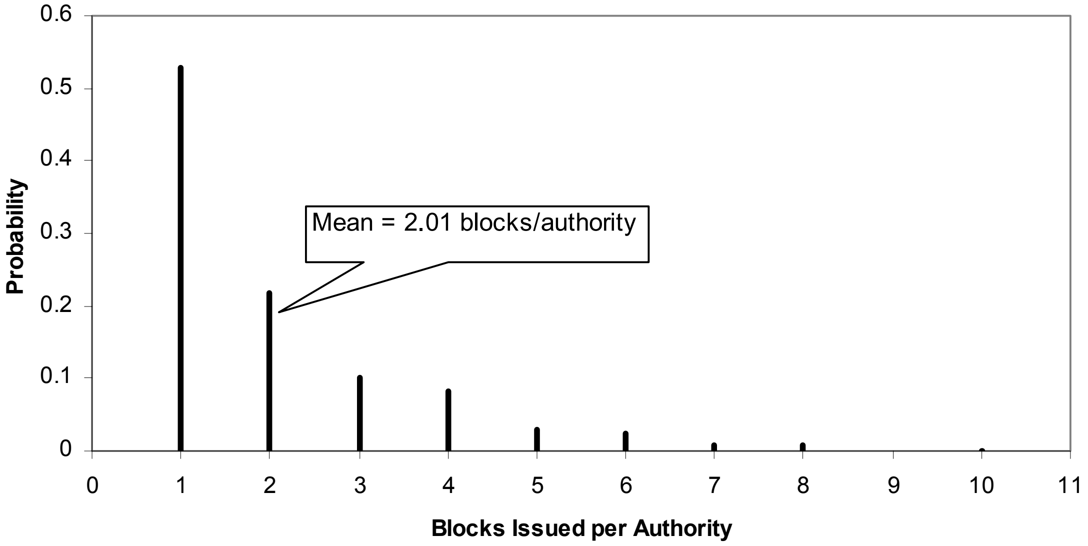


Figure 4. Distribution of Number of Blocks Issued per Authority

Using this distribution, the mean number of authority boundaries traversed per trip along the test territory was 9.43. There were 19 blocks between the ends, not including the yards (outside the scope of this analysis). The corresponding annual number of authority boundaries traversed was:

9.43 boundaries x 273 trains in 2 weeks x 26 2-week periods in 1 year = 66,940 boundaries/year.

Dividing the distribution of authority exceedances (/10 years) by 669,400 boundaries (the equivalent 10 year rate) yielded a distribution with a mean rate of 6.7 x 10^{-6} per authority boundary.

Quantification

The CSXT-wide analysis led to an estimate of the exceedance rate to be a distribution with a mean of 3.14 x 10^{-7} events per train-mile. The territory-specific estimate was a distribution with a mean of 5.26 x 10^{-7} per train mile. This difference of a factor of two was considered not significant, given the number of assumptions used to generate them.

The ASCAP analysis modeled the exceedance rate per block, not per train-mile, in its estimates. Given that there are 19 blocks along the test territory of 120.5 miles, the average block length is 6.3 miles. Therefore, the mean exceedance rate per block using the CSXT experience is 1.99 x 10^{-6} events per block, and using the territory experience is 3.34 x 10^{-6} per block.

However, to select between these two results, their distributions were compared. The comparison is shown in Figure 5.[11]

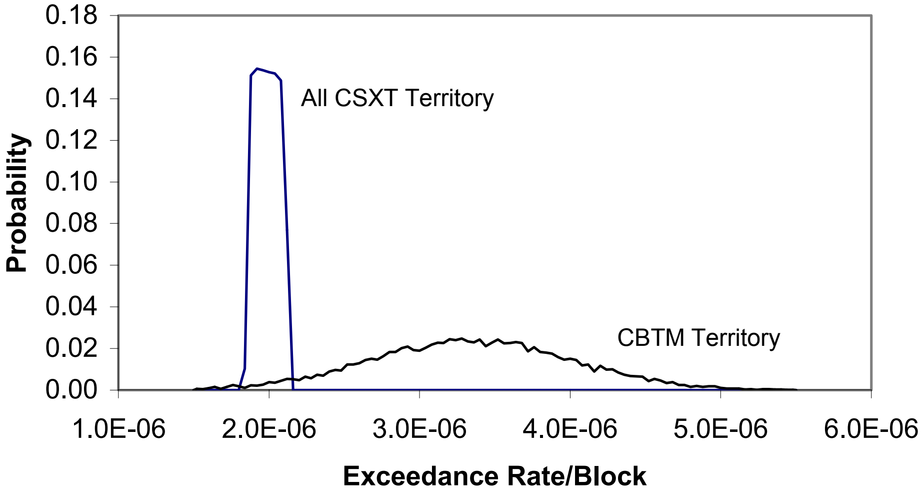


Figure 5. Exceedance Rate per Block for all CSXT Territory and CBTM only Territory

When the distributions are compared, they overlap, with the CBTM territory results distribution extending past the 'all CSXT' distribution. Based on this comparison, the workshop participants preferred the CBTM territory result since its mean is slightly more conservative, and its distribution enclosed that of the CSXT-wide analysis. Therefore, the distribution for use in

[11] The figure is a result of the discrete event simulation used in Crystal Ball, and therefore the results appear as histograms.

ASCAP for the probability of a train crew to exceed its limit of authority can be approximated by a normal distribution having a mean value of 3.3 x 10^{-6} per block boundary and a standard deviation of 6.8 x 10^{-7}.

As mentioned in Section 3.4.3, an alternative set of summary data for the movements through the CBTM territory in a different two-week period identified 400 total train movements, compared with 273 movements in the data described above. To investigate the effects of this alternative data set, the above calculation was repeated using to 400 movements. The result was a lower mean probability of the train crew exceeded their limit of authority (2.3 x10^{-6} per block, versus 3.3 x 10^{-6} per block [above]). This difference (30% reduction), while not trivial, is still within the range of the uncertainty calculated for the original data and is still larger than the value calculated using the CSXT-wide data. Using the alternative data would not significantly alter the values recommended to ASCAP and lie within the range of uncertainty.

3.5.2 Summary of other CBTM related Analyses

This section provides a summary of the qualitative analysis and quantification for each of the remaining events that were quantified as part of the quantification workshop. Details of the qualitative discussions and quantitative analysis are provided in Appendix C.

Dispatcher-caused Boundary Exceedances

Qualitative Discussion

Workshop attendees indicated that the dispatcher can create conditions where the train crew believes they have a valid authority (based on the verbal communications with the dispatcher) but they are unprotected by the CAD system, which could allow an authority to be issued to another train. Examples of how this could occur include:

- Errors related to use of the CAD system
- Train misrouting
- Radio miscommunications

Workload and problems with radio communication were identified as the most important influences on the likelihood of dispatcher-caused exceedances:

Quantification

The workshop attendees agreed that the most immediately relevant databases were the dispatcher disciplinary data and the CSXT operating data from FRA.

The mean rate for dispatcher-caused exceedances was calculated to be 3.5 x 10^{-6} exceedances/block boundary. Based on the Crystal Ball analysis, the best fit for the resulting distribution is a Gamma distribution, with a location parameter of –2.02 x 10^{-6}, a scale parameter of 5.12 x 10^{-8}, and a shape parameter of 7.73 x 10^{1}. Figure 6 shows the resulting histogram and the associated best-fit continuous distribution.

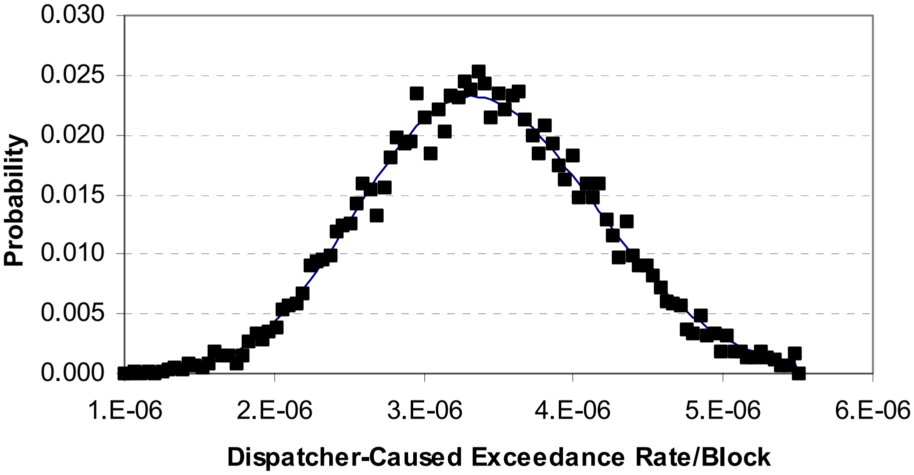


Figure 6. Histogram and Best-Fit Distribution for Dispatcher-Caused Exceedances

Overspeeding Events

Qualitative Discussion

Two types of over-speeding events were considered: Permanent and temporary speed restrictions.

The most important influence for overspeeding in a permanently restricted section is experience and knowledge of the territory. Most inadvertent overspeeding events were described by the workshop participants as occurring in the first 2 years of experience. Other contributors include train make-up (e.g., brake profile) and weather conditions that can impact braking and visual cues (e.g., icing can affect ability to brake, fog can affect ability to see signs, etc).

In the case of temporary speed restrictions a variety of influences for overspeeding were mentioned. In addition to the factors that apply to permanent speed restrictions, some factors are unique to temporary speed restrictions. These include the fact that there can be many (upwards of hundreds) temporary speed restrictions active at any given time. In most cases, these temporary restrictions are printed in the train messages and train bulletins. Some temporary speed restrictions do not appear in the printed train messages and bulletins. These are communicated to the train crew en-route by the dispatcher over the radio. Train crews can miss a temporary speed restriction (e.g., inadvertently skip over it in train message or train bulletin) due to the large number of temporary speed restrictions they have to manually track. The temporary speed restrictions communicated via radio are subject to additional opportunity for error (e.g., hearing or writing).

Quantification

A total of 56 overspeeding events were identified in the train crew disciplinary database over a 4-year period, corresponding to an annual rate of 14 per year, assessed for DTC operations. These overspeeding events were sufficiently serious to lead to the engineer being decertified by CSXT. The workshop participants described the most likely range for underreporting of overspeeding events to be in the range from 2 to 4 times, with 3 being the most likely.

The calculated rate for exceedances per restriction is a distribution with a mean of 4.6 x 10^{-6} exceedances per speed restriction. This can be approximated by a Beta distribution, with the following parameters: α = 1.22 x 10^{1}, β = 1.73 x 10^{1}, scale = 1.1 x 10^{-5}. Figure 7 shows the summary histogram from the Crystal Ball simulation and the fitted Beta distribution.

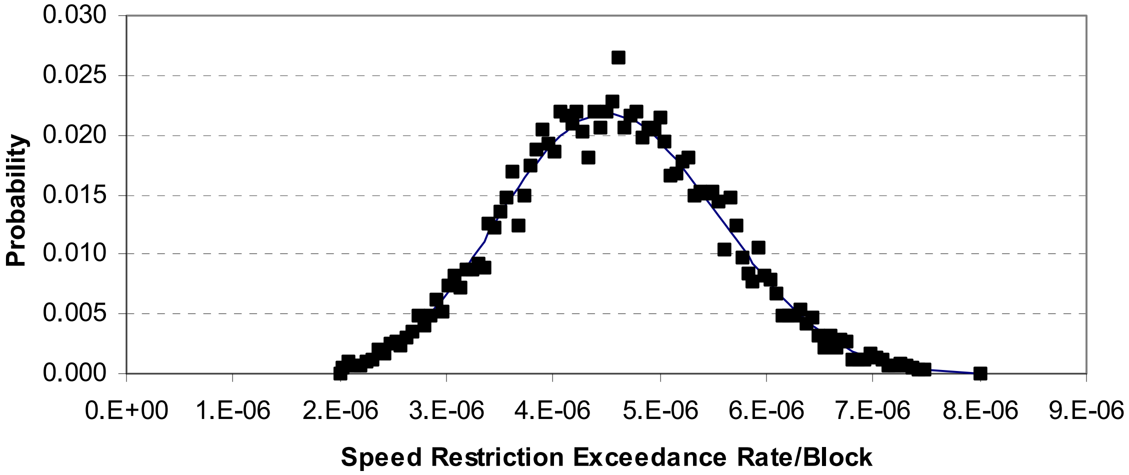


Figure 7. Histogram and Best-Fit Distribution for Speeding Events

<u>Switches</u>

This analysis is presented in two parts: the likelihood of a manual switch being left in the wrong position, and the likelihood of a train running over a mis-positioned switch.

Manual switch in wrong position

Qualitative Discussion

Workshop participants discussed at length some of the factors that could lead someone to leave a switch in the wrong position. They mentioned that, while it was not the established operating procedure, there appears to be evolving an acceptance of the practice of allowing train crews to re-align switches for each other. conductors will only leave a switch in the reverse position if they get positive confirmation from the other train crew (via radio communication) that they will re-align the switch for them.

One reason for this evolving practice is that with the change to a two-person crew and no caboose, the conductor may need to walk a long distance (as much as 150 car lengths) to get back to the cab after re-aligning the switch. This is time consuming, delaying movement of the train. It can be particularly difficult and time consuming in cases of bad weather or poor paths for walking. As a result, sometimes a train crew in a siding will leave a switch behind in the reverse position if they have verbal agreement with the conductor of another train that they will re-align the switch (based on radio communication between the conductors of the two trains.)

Generally, the switch will be correctly re-aligned by the conductor on the next train. However, factors that can contribute to failure to correctly re-align the switch include:

- Miscommunication (e.g., due to poor radio reception)
- Distraction – leading to forgetting
- Changing plans (e.g., there is supposed to be a following train, but then plans change).

Quantification

None of the databases available for the workshop provided useful data for this analysis. Therefore, the primary inputs were from the workshop participants, particularly those with current relevant experience (the engineers and conductors on the route).

The distribution of the likelihood of a switch being in the wrong position at the time a train approaches has a mean value of 1.3 x 10^{-4} per train. The histogram and the associated best-fit triangular distribution are shown in Figure 8. The triangular distribution has a minimum of 2.8 x 10^{-5}, a maximum of 2.8 x 10^{-4}, and a most likely value of 5.7 x 10^{-5}.

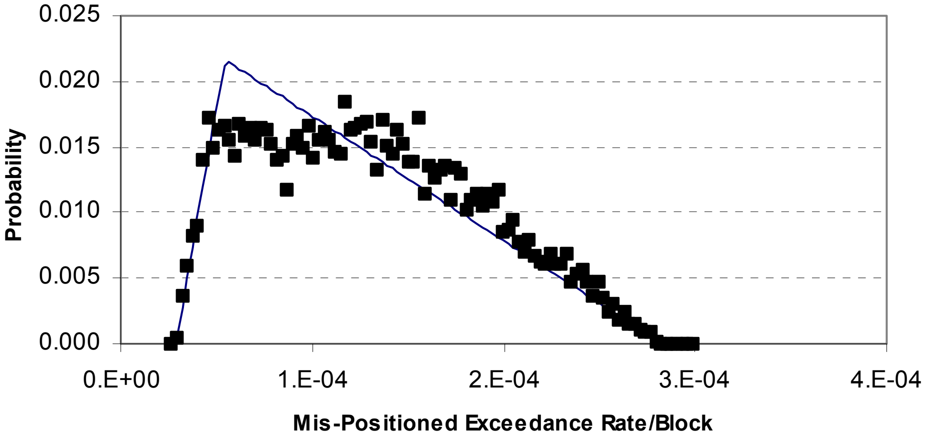


Figure 8. Histogram and Best-Fit Distribution for Mis-Positioned Switches

Train runs over wrongly positioned switch

Qualitative Discussion

Workshop participants indicated that the major factor determining whether a train would run over a wrongly positioned switch was the ability of the crew to see the target in time to stop. At track speed, the ability to stop in time was felt to depend on factors such as train speed, terrain grade, trainload, and visibility. If traveling at slow speed (less than 10 mph), the likelihood of failing to stop was considered very low.

Quantification

Of the 10 manually positioned switches along the length of the route, crews stated that (because of the visibility of the specific switch targets) they would not be able to observe the state of 7 switches when traveling southbound and 6 switches when northbound in sufficient time to stop before running over the switches when traveling at track speed. Of the 3 southbound and 4 northbound switches where at least the potential existed for stopping, the distribution of the probability of being able to stop in time when traveling at track speed had a mean value of 0.22. It was represented by a triangular distribution whose minimum is 0.15, maximum is 0.3, and with a most likely value of 0.2. No simulation runs were performed; this distribution is a direct input to ASCAP.

If traveling at slow speed (less than 10 mph), the likelihood of failing to stop was considered very low (1 in 10,000).

Work Zones

Qualitative Discussion

The scope of the analysis covered work performed under Rule 707 – preplanned work authority that appears in the train bulletin. The error of concern was entering the work zone during the restricted times (as specified in the train bulletin) without authority, or being in the area at the start of the specified time.

Participants mentioned a number of factors that could contribute to entering a work zone without authority. These included:

- Communication errors (e.g., due to poor radio quality)
- Crew thinks it will clear the affected DTC Block before the work-zone is activated
- Misinterpreting location of the work zone (for less experienced crews)
- Intersecting lines. If a work zone is on one line sometimes trains that are passing through on an intersecting line may not be aware of the work zone.
- Missed red (stop) board (The same issues and factors as in the case of missing a block authority boundary sign.)

Workshop participants had little direct knowledge of cases of work zone entry without authority in the Augusta to Spartanburg territory.

It was brought up that the employees who work on the Augusta to Spartanburg territory tend to have many years of experience. As a result, the error rate may be low for train crews and roadway workers, but that in other territories with less experienced personnel (or in the future in this territory) there may be higher error rates.

It was mentioned that there might be under-reporting of instances of going through work zones as well.

Quantification

Data necessary for this event were not available. Workshop attendees suggested using the same fraction as for exceeding DTC block authority

CBTM Applications

Three conditions were analyzed for the use of the CBTM system:

1. The crew fails to gain control of the locomotive/train following indication of a warning before the penalty brake is applied
2. Train crew over-relies on CBTM (a complacency effect)

3. The train crew enters incorrect consist information into the CBTM system

These three events were selected for quantification based on requirements defined by the PRA, as well as results of the qualitative analysis that was conducted prior to and during the quantification workshop that suggested that these events were situations of potential concern.

The workshop attendees agreed that there was insufficient experience with the CBTM system to confidently project its potential impact on human performance. The local CSXT locomotive engineers and conductors indicated that while they had the most experience with CBTM, they have only had the opportunity to operate CBTM equipped trains a couple of times each. Further, the field-tested version of the CBTM prototype was expected to improve substantially prior to actual implementation. Consequently, experience with the CBTM prototype was not expected to be representative of performance of the final production system.

Given that opinion represented the consensus position at the workshop, participants recommended performing sensitivity studies to explore how different assumptions about the impact of CBTM on human performance would affect the results of the CBTM case.

The results for each of the three individual CBTM issues discussed at the workshop are summarized below. In some cases numeric probability estimates were elicited from the workshop participants. These estimates are presented along with the assumptions that served as a basis for the probability estimates. These probability estimates are recommended as starting points for sensitivity analyses.

The crew fails to respond to a warning before the penalty brake is applied

Qualitative Discussion

In this case, a penalty brake occurs that may have been avoidable. There are several reasons why the crew may fail to prevent a penalty brake. In the current trials, the locomotive engineers reported that the time to respond seemed short (CSXT planned to re-examine the response time between the warning and penalty brake application at the end of the trials). Second, the CBTM system did not recognize when dynamic braking was applied; it only recognized application of the air brakes. Therefore, the locomotive engineer may use one braking system without CBTM "awareness."

For quantification purposes, the workshop participants decided to assume that the production system would include design changes to avoid some of the limitations of the current prototype CBTM. Specifically, the workshop participants recognized that the CSXT would consider revising the braking algorithm following the test period.

Quantification

Given the limited testing experience with CBTM, and uncertainty in the final design, the consensus of the participants was that the range of probabilities for failing to respond in time to a CBTM warning to prevent a penalty brake in the final design was in the range of 0.1 to 0.01. This was represented by a lognormal distribution with its 5th percentile value of 0.01, and its 95th percentile value of 0.1 and is truncated at 1.0. The mean value of this distribution is 0.04. Again, no simulation runs were required: this distribution is a direct input to ASCAP.

Given the uncertainty expressed by workshop participants, the facilitators recommended performing sensitivity analyses using the ASCAP simulation to explore what the impact would be if the probability of failing to respond in time to a CBTM warning to prevent a penalty brake was higher than 10%.

The train crew over-relies on CBTM (a complacency effect).

Qualitative Discussion

This case relates to the potentially negative effects of over-reliance on CBTM (i.e., complacency). Might the train crew grow to rely on CBTM to catch problems such as over-speeding or entering a block or work zone without authority? Would the train crew become less vigilant in preventing such occurrences from arising? Such complacency can happen as an unintended consequence of using new technologies. The safety consequences of such over-reliance emerge when the CBTM system fails. If the train crew comes to rely on CBTM, and the system fails, then the event for which CBTM would ordinarily provide backup protection (i.e., over-speeding, exceeding an authority, unauthorized entry into a work zone, or over-running a protected switch) would be more likely to occur.

At the workshop, participants agreed that the experienced crews on the Augusta-Spartanburg run would operate under the philosophy that CBTM should never actuate and that their experience will enable them to avoid nearly all warnings. They may operate near the limits of the system's operating envelope, but act early enough to avoid CBTM warnings. Under this assumption, complacency will not be an issue since there is no reliance on and no regular occurrence of CBTM warnings. Therefore, for current operation of the Augusta-Spartanburg run, there is no change in operator error probabilities, given CBTM is in use but having failed.

Note that, for different crews, territories, railroads, and operating philosophies, this condition may change. The workshop participants stressed that crews on the Augusta-Spartanburg territory were particularly well experienced, and that less experienced crews might not be expected to perform in the same way. Experience in other industries suggests that as systems like CBTM become "the norm", the people using the system will adapt their modes of working around that system. Unless management and the crews maintain a strong focus on the system being used only as an overlay system, it is possible that crews will become more likely to rely on it. If so, crew failures to act may be more likely with CBTM than without it.

Quantification

For current operations in the Augusta-Spartanburg territory, there is no change in operator error probabilities from the base case analyzed (operations without CBTM). The probabilities of crew failures (e.g., exceeding block authority, crossing an improperly positioned switch, entering work zones without authority or overspeeding) would be the same with CBTM, as in the base case.

However, for reasons discussed in the last section, other railroads, crews, and changing operating philosophy, crew failures to act may be more likely with CBTM than without it. Therefore, the ASCAP (or any future similar) analysis should perform sensitivity studies to explore the potential impact of different assumed levels of complacency on mishap rates. The effect of complacency on human performance can be modeled by increasing the human error rates in the CBTM case relative to the base case human error rates. The potential effect of different levels of complacency could be explored by multiplying the base case human error rates by factors of 2, 5, and 10 to obtain the human error rates in the CBTM case. Changes in mishap rates output by the ASCAP model as a function of different assumed levels of complacency would provide CSXT and the FRA with information as how important it is to ensure that complacency does not occur.

The crew enters incorrect consist information into the CBTM system

Qualitative Discussion

Discussions of this case centered around two situations:

- Intentionally entering incorrect consist information (e.g., to manipulate the CBTM braking profile)
- Unintentional errors (e.g., a data entry error by the crew, a failure to remember to update information in CBTM when a change is made to the consist, or an error in the consist description provided in the paper work given to the train crew)

Quantification

Workshop participants felt that while there was a potential for incorrect consist entry into CBTM, there was insufficient experience to estimate the probability of incorrect entry. For this reason, no quantification of the likelihood of incorrect consist entry was provided. Participants felt that intentionally entering wrong consist information was unlikely because management is likely to impose disciplinary action for deliberate manipulation of the consist information.

4. LESSONS LEARNED

A number of challenges were confronted in performing the human reliability analysis for the CBTM case. These challenges included how to define the appropriate level of decomposition in modeling human performance, how to illicit and integrate the perspective of the multiple stakeholders in the human reliability analysis and quantification process, how to use the results of the qualitative analysis to inform the human reliability quantification process and how to estimate the impact of a new technology on human reliability in a case where the technology was still under development and the user experience base was limited.

These challenges are not unique to the CBTM case. Similar challenges are likely to be faced by any human factors quantification project. In this section, we summarize the lessons that were learned from the CBTM case that have broad applicability.

This section provides generic recommendations for how to conduct a human reliability analysis based on the lessons learned from the CBTM case. The specific challenges that arose in the CBTM case and how they were addressed illustrate the generic recommendations.

4.1 Establish Appropriate Level of Decomposition for Modeling Human Reliability

One of the first tasks in performing a human reliability analysis is to determine the appropriate level of decomposition for modeling human behavior. In developing a model of railroad operations, human failure events can be defined at a high level such as ‘train fails to stop at the limit of authority due to train crew error’ or at a more detailed level, with multiple sub-elements each corresponding to a more specific type of human error. For example the human failure event ‘train fails to stop at the limit of authority due to train crew error’ might be further decomposed to include more specific ‘human failure events’ such as:

- ‘Train crew fails to recognize cues to stop at the limit of authority’,
- ‘Train crew fails to stop because they incorrectly believe they have authority to proceed due to communication error’, and
- ‘Train crew fails to stop at the limit of authority due to misjudgment of required braking distance (given locomotive engine characteristics, consist, and environmental conditions)’.

Typically PRA models can accommodate human reliability estimates at multiple levels of decomposition. The appropriate level of decomposition at which to model human performance will depend on several factors.

One factor is the availability of credible data sources to support modeling at the different levels of decomposition. It is important to establish a level of decomposition in the human reliability model that is consistent with the available data as well as the ability of domain experts to provide estimates of human reliability. Domain practitioners are most comfortable providing estimates at a grain of analysis for which data exist that they can use as a point of comparison. The quantification workshop provides a vehicle to integrate 'hard data' and 'expert judgment'. This comparison is most easily accomplished when the data and expert judgments are conducted at the same level of decomposition.

A second factor is whether human performance can be modeled accurately at a particular level without making overly simplistic or artificial assumptions. In establishing the level of decomposition at which to model human performance, it is important to consider the possibility of dependencies between the human activities modeled. For example, in the railroad industry 'readback' is typically used as a way to catch and correct communication errors. However, some factors that impact the probability of making a communication error (e.g., a noisy radio channel that causes the train crew to mishear the block limit of authority that the dispatcher gave) can also affect the probability that the error will NOT be caught during the readback (e.g., the fact that the radio channel is noisy will increase the probability that the dispatcher will mishear the readback and think that the train crew repeated back the block authority that the dispatcher gave). If the dependency is ignored or forgotten in the HRA model, it is possible to underestimate the overall probability of failure by treating the activities as if they are independent during quantification.

A third consideration in defining the appropriate level of decomposition, is whether the distinctions made at the more detailed level of analysis are important for the question(s) being addressed by the HRA analysis. In evaluating the introduction of a new technology, the question is what level of decomposition is required to evaluate the impact of the technology on human performance. For example, CBTM is designed to prevent train crews from exceeding their block of authority, independent of the reason for their action. CBTM is insensitive to whether the crew is about to exceed their block of authority because of distraction, a communication error, or a misjudgment of required braking distance. As a consequence, there is no particular need to model human performance at a detailed level of decomposition to assess its impact on safety. In contrast, there may be other technologies that selectively impact particular elements of human performance, such as communication. In those cases, it may be necessary to model the details of human communication in railroad operations in order to establish the benefits of the new technology. An example is improved communication technologies (e.g., improved analog radios or digital communication devices), where the benefits of the technology can best be established by explicitly modeling the communication that occurs between dispatchers, train crews, and roadway workers.

CBTM Example

In the CBTM example, human error events were defined at a high-level of granularity. The primary reason human reliability was not modeled at a more detailed level of decomposition (e.g., recognition errors, communication errors, braking errors) is that the available data did not support human error estimation at that detailed a level of decomposition. There was no readily available quantitative data from the railroad industry from which to estimate error values at a detailed level of decomposition. There was some data of potential relevance from related industries (e.g., data on communication errors between pilots and air traffic controllers in aviation). However, when railroad experts at the preliminary quantification workshop were presented with this data (e.g., Cardosi, 1993; Cardosi, Brett, & Han, 1996), the railroad experts indicated that the parallels between railroad dispatching and air traffic control operations were not sufficiently similar to allow confident extrapolation from the aviation domain to the railroad domain. Similarly, railroad experts at the quantification workshop were unable to provide human reliability estimates at a more detailed level of decomposition (e.g., probability of a communication error).

The experts gathered at the quantification workshop were able to draw upon their own experiences to estimate high level failure events by recalling how often it happened to them or to others that they heard about (e.g., how often they themselves exceeded a limit of authority in X years, or heard about others who did so over the past X years). Participants were readily able to recall severe or unusual situations that occurred to them or that they heard about (e.g., accidents, near-misses).

The experts at the quantification workshop were uncomfortable providing human performance failures at a detailed level of decomposition (e.g., recognition failures; communication failures). The existence and impact of these more elemental human performance failures were discussed as part of the process of generating probability estimates for the higher level human failure events (e.g., communication failures can lead a train crew to believe they have authority to proceed into the next block when in fact the block authority was not issued by the dispatcher) but the workshop participants did not quantify human error at this more fine grained level of analysis.

The domain practitioners were comfortable providing estimates at a grain of analysis for which some data existed that they could use as a point of comparison. As discussed in Section 3.4 the objective of the quantification workshop was to provide a vehicle to integrate operational data and the experience of operational experts. This is most easily accomplished when the data and expert judgments are conducted at the same level of decomposition. The workshop participants were very comfortable with that approach.

A second consideration in determining the level of decomposition for human error modeling in the CBTM case was the extent to which human performance could be accurately modeled at a particular level, without making overly simplistic or artificial assumptions.

One case in point was communication between train crews and dispatchers. In the simplest case the train crew initiates a call to the dispatcher when they come to the end of a block authority to release the block they have just completed and request block authority for the next block. A simple model can be created to capture this pattern of train crew – dispatcher communication. However, interviews and observations of train crews and dispatchers indicated that, while many communications between train crews and dispatchers conformed to this simple case, there were other communication patterns that occurred as well. For example, a dispatcher will frequently provide authority for several blocks at once (See Figure 4). There were also documented cases where the dispatcher asked the train crew to release blocks in groups rather than call in to release each block as it was passed. Finally, dispatchers and train crews indicated that the dispatcher would initiate a call to the train crew to either request release of several blocks or provide block authority for the next several blocks.

These varied patterns of communication indicated that it would be more difficult to accurately model train-crew dispatcher communication than the simple case suggested, and was a factor in the decision to model human reliability at a high level of decomposition that did not require modeling the details of communication.

4.2 Insure Broad Participation by Stakeholders

It is important to insure broad stakeholder participation in the selection of participants in both the qualitative analysis and the quantification workshop so that the different stakeholder communities will perceive the process and results as credible. The process of generating the human reliability values relies heavily on input from domain experts. This is true for both the early qualitative analysis phase, where observations and interviews of domain practitioners are conducted, as well as the later quantification workshop phase, where experts get together to generate reliability estimates based on a consensus process

If the individuals selected for participation do not represent a broad range of perspectives (e.g., management, labor, vendors, regulators) then there is a danger that the results will be unrepresentative. Therefore, it is important to insure broad stakeholder participation in the selection of participants in both the qualitative analysis and the quantification workshop to insure that the different stakeholder communities will perceive the process and results as credible.

It is particularly important to have broad stakeholder participation in the quantification workshop. The quantification workshop represents the point at which the available quantitative and qualitative evidence is presented and evaluated. It is important to insure that all the various stakeholders have an opportunity to participate in the process. Broad stakeholder participation is a critical element in the credibility and acceptance of the human reliability quantification process and final product.

CBTM Example

In the CBTM demonstration case, the RSAC working group provided access to a broad range of stakeholders. Broad participation in the qualitative analysis was achieved by soliciting help of local CSXT management and local labor representatives in identifying the individuals to observe and interview. Interviews were held with the individuals directly involved in the work (locomotive engineers, conductors, dispatchers) as well as individuals in the management chain, and trainers. These individuals had direct experience with the CSXT territory in question and with the CBTM prototype. RSAC members provided broader perspective on potential issues of concern to consider as part of the qualitative analysis. RSAC meetings provided the vehicle for obtaining their input.

A concerted effort was made to solicit broad stakeholder representation in the quantification workshop for the CBTM demonstration case. Labor, railroad management and regulatory members of the RSAC committee were invited to participate in the workshop and/or send representatives. Labor members of the RSAC were requested to help identify and solicit local and national representatives to participate in the workshop. In total 30 people participated in the workshop including four railroad representatives, thirteen local and national labor representatives (including locomotive engineers, conductors, dispatchers, and roadway workers), and six FRA representatives.

In addition, the RSAC committee provided a review of the results of the analysis, including the workshop and its inputs. While brief, this review did discuss the scope of many of the analyses, the data used, and the conclusions from the results. This provided an opportunity for RSAC members and other stakeholders to evaluate and discuss the process and the results.

4.3 Select Data Sources and Their Uses

The selection of data sources and their uses as a basis for quantifying failure probabilities is a critical issue in performing human reliability analyses in cases where operational experience is available and relevant to the scope of the study. However, as discussed elsewhere in this report (particularly Section 2.3.4), an exact match between the scope of the analysis and the contents of the databases is rare. Therefore, the analysis must compare the databases against the scope of the study to identify the most relevant databases and what kinds of adjustments may be necessary. This section discusses how to approach the matching process and the process for making adjustments.

As described in the example analysis, crew exceedance of their authority (Section 3.4), more than one database may be relevant to a particular analysis. It is not necessary to limit the analysis to using only one of the available databases, nor is it appropriate to "force-fit" data from multiple databases into a single source before examining the results of using the databases separately.

In considering the relevance of a particular database, it is important to consider how the scope of the database and the scope of the analysis compare. This comparison needs to consider the formal and the informal rules by which data are included in the database. For example, the incident database compiled by FRA is based on reports submitted by railroads and includes a range of potential cause codes, including those associated with human errors, as discussed in Section 3.4.2. However, there is no certainty that the cause identified is uniquely and unambiguously correct. Most accidents are the result of multiple causes and therefore "the cause" (when using databases for which only one cause can be identified) will always be incomplete, and may be biased by a desire to underemphasize causes that are "politically sensitive". Often, even where databases have fields for narrative text, there is often very limited information provided in them, though the analyst should always check for any useful information there.

It is important that the analyst get background information from as many sources as they can about any known limitations or biases in the databases being considered for use so that adjustments can be considered in the elicitation process. For example, if the analysis were to be performed on human caused speeding events, then a naïve analysis would just consider how many events were identified as being caused by human errors in the cause codes of the FRA incident database. A better approach would be to ask experienced data analysts who use the same database what other kinds of cause codes could be used to describe overspeeding events that might be human-caused. In the elicitation process, the knowledgeable participants can be asked to assess what range of events under these other cause codes could be human-caused.

Section 2.3.4 discusses the general approach to filtering and scaling the databases. Filtering is the process for removing events from the database where they are not relevant to the scope of the study. Scaling is the process for making adjustments for events that may be missing from the database but are within the scope of the study. Filtering involves judgments about what is within the scope of the analysis. The scope may need to be clarified within the scope of the overall risk analysis. A common example is the degree to which rule violations are represented by events in the database. Would a minimal rule violation be considered an "event" in the PRA? There is no uniquely correct answer. The people conducting HRA and PRA tasks must resolve this through discussion. Other factors of concern in the filtering process include identifying events that cannot occur in the system under study. For example, events occurring in signal territory may need to be excluded from DTC operations.

In the case of scaling, it is necessary to identify potential gaps in the databases being used. In some cases, a database may have criteria for events to be included, such as the damage or casualty criteria for the FRA database. Incidents involving no injuries or losses less than $6,500 are not reportable to FRA, though they may be relevant to the events being modeled. Potential disciplinary events that are not detectable other than by self-reporting by the dispatcher or crew involved are likely to be significantly under-reported. Any expert elicitation workshop must include attendees who can identify the possible gaps in databases and estimate ranges of scaling adjustments to the data.

In making the necessary adjustments for the filtering and the scaling, it is strongly recommended that ranges be estimated for the adjustments, rather than trying to identify point (single) values. As discussed in Section 2.3.5, the elicitation process should aim at identifying the consensus of the group in terms of the possible range of parameters. There is almost never a single "correct" parametric adjustment, and attempting to reach such a single value is more often counterproductive because of differences between experts that may lead to disagreements and disputes of each other's "validity" in trying to reach consensus on a single value, whereas agreement on a range can often be accomplished more readily.

CBTM Example

In the analysis of the crew exceedances, described in Section 3.5.1, the analysis proceeded using all the available databases and the results were based on a comparison of the analyses using each database. In that case, using the broader of the two final distributions encompassed the second (narrower) distribution, and was considered more realistic in its portrayal of the uncertainties.

In considering some of the data represented in the crew disciplinary database associated with exceedances, it was reported by the workshop participants that some events could be as small as a matter of inches when the database records engineer de-certifications during efficiency testing. In addition, other data suggested that a significant fraction (about 40%) of exceedances during normal operations (i.e., not during testing) are less than 100 feet. In this analysis, all exceedance events in the database were included in the analysis since the criterion for significant exceedances was being developed as a later part of the analysis of potential consequences in the ASCAP PRA study.

In terms of making scaling adjustments, the workshop participants relied on the knowledge of the experienced train crews to estimate how often they thought such exceedances really took place, compared with the predicted rate using the data from the database. It was noticeable that the estimates for the territory-related data, the events with which the crews were familiar, represented the broader range and was used in the final analysis. The use of experienced crews as a source of data is generally the preferred source.

4.4 Use Qualitative Information to Guide Quantitative Analysis

Qualitative analyses contribute substantially to the quality of the outputs of the quantification workshop. The qualitative analysis preceding the quantification workshop provides the background needed to structure and lead the workshop. It allows workshop facilitators to raise possible causes and contributors to error for discussion by workshop attendees that might otherwise be missed, thus insuring more thorough discussion of issues by workshop participants prior to quantification.

The qualitative discussions that occur prior to quantification help to ground the workshop participants, providing them with a common understanding of the causes and contributors to errors, and a concrete basis for estimating frequency of occurrence. A qualitative discussion of the causes and contributors to a type of error provides participants with a concrete understanding of the variety of ways that an error can arise. For example, in the case of exceeding speed limits, factors such as temporary speed restrictions and missing or obscured signs create conditions that make it more difficult for train crews to recognize the need to reduce speed. In those cases, the kinds of cues that are normally present to remind the crews that they need to reduce speed (e.g., familiar landmarks, clear sign posts, well-practiced routine) are not available. As a result, errors are more likely to occur. Decomposing error events into the different ways they can occur can provide additional leverage in trying to estimate error frequencies. In the case of exceeding speed limits for example, the participants can estimate what proportion of time a sign is likely to be missing or obscured, and then estimate the likelihood of failing to reduce speed given that the sign is missing.

CBTM Example

In the CBTM demonstration case, a background package was prepared for each of the human error events to be quantified. This background material included a list of possible causes and contributors for the error events to be quantified. The list was derived from the interviews and observations conducted prior to the workshop.

The background package served two purposes:

1. It provided the workshop facilitators the necessary background to moderate workshop discussions and rapidly assimilate the points being made by the expert participants at the workshop.
2. It provided a ‘checklist’ of issues that the workshop facilitators used to guide discussions and insure that all potentially important causes or contributors to error were considered and discussed.

Since the point of the quantification workshop was to elicit the causes of human failure events and associated probabilities from participants, the list of individual unsafe actions and contributors that can lead to the error events was never explicitly presented. The workshop facilitators used it as a framework to guide the qualitative discussions that preceded quantification of error events. In almost all cases, the workshop participants brought up all the items included in the background package without being prompted. However, if a particular issue did not come up during the qualitative discussion, then the facilitator raised the topic. The facilitator mentioned that an item had been brought up in earlier interviews and asked the workshop participants whether they perceived it as important. This helped reduce the chance that a significant contributor to error would be overlooked.

Qualitative discussions occurred prior to quantification of each human error event. This served to provide all participants with a common understanding of the events and factors being analyzed. The scope of the human error event under consideration, and the different causes and contributors to that error were discussed. Discussing the different ways that errors could occur provided additional leverage in trying to estimate error frequencies.

4.5 Project Impact of New Technology Given Limited Experience

One of the challenges in doing a human reliability analysis in support of a performance-based evaluation of a new technology, is the need to project what human reliability values are likely to be once the new technology is introduced. This is challenging because much of the information required to make informed estimates is often unavailable.

A risk-based evaluation is likely to be performed early in the technology development process. Typically, the risk assessment is conducted at a point in the system development where only early prototypes are available and few if any field tests have been conducted. As a result, there is likely to be little if any quantitative data from which to estimate the potential impact of the new technology on human performance. It is also difficult to utilize expert judgment techniques to estimate human error because target users are likely to have limited experience in using the system and therefore limited ability to predict how the system will impact human performance and human error.

Moreover, the early prototypes may not accurately reflect the ultimate system that is implemented. For example, often an early prototype will exhibit human factors problems that could negatively impact human performance and human reliability if they are not corrected prior to final implementation. However there is high likelihood, that these human factors problems will be corrected prior to final implementation. Conversely, a prototype may exhibit positive features that contribute to high human reliability that may ultimately not be manifest in the final production system (e.g., because those features may prove to be too expensive to implement). This creates a dilemma, should the human error estimates be generated based on the existing prototype, on the vendor’s vision of what the ultimate system to be implemented will be like, or on a worst-case scenario?

Finally, even if the prototype accurately reflects the properties of the final system to be implemented, there are likely to be many operational details with respect to how the system will be used that have not been finalized. Examples include: (1) what training will be provided to railroad personnel? (2) What policies and procedures will be put in place, and (3) whether the new technology will be implemented throughout the railroad (e.g., on every locomotive that travels through a territory) or in a more limited way. These operational details are likely to profoundly impact human reliability.

Several strategies can be employed to overcome these sources of uncertainty regarding how the final implemented system will impact human performance. One approach is to make specific assumptions about the system that will ultimately be implemented, and generate human reliability estimates given those specific assumptions. In that case it is important that the assumptions made are documented, and that the system that is eventually implemented is shown to meet those assumptions. Another approach is to perform sensitivity studies to explore the safety consequences of alternative assumptions about the new technology and its impact on human performance.

Whatever approach is used to project the potential impact of a new technology on human performance, person-in-the-loop[12] tests should be performed prior to final implementation of the production system to establish that the system-as-built accords with the assumptions that were made in the safety case analyses. Typically field tests of train control systems are designed to evaluate that the hardware and software systems meet predefined performance criteria. Person-in-the-loop test are tests that evaluate not only the hardware and software, but also the ability of

[12] “Person-in-the-loop” refers to the concept in the modeling, design, and testing of systems, where the nature of human performance, including the associated limits of information processing (memory, speed, accuracy, etc.) and error potential are fully taken into account, especially by testing with “real people” under realistic test conditions, such as in a high fidelity simulator or in the actual work environment.

the people in the system (e.g., the train crew) to perform the tasks expected of them by the system design (e.g., to detect warnings in time to take appropriate action, to perform control actions correctly within the available time window). Inevitably, any simulation-based analysis depends on assumptions made about how the new technology will perform. One way to use a simulation tool such as ASCAP is to ask the following question: If a technology functions in the way that its developers expect it to (such as its reliability, accuracy, braking strategies, number of false alarms, and impact on human performance), will it reduce risk?

In submitting a final safety case, evidence needs to be provided to show that the proposed technology does in fact meet the criteria that were assumed in the simulation-based analyses. While the question of whether a system that operates in a postulated way will improve safety can be addressed through a simulation model, the question of whether the system that is ultimately implemented actually operates the way it was postulated to operate in the simulation model can best be established via empirical 'person-in-the-loop' validation tests that evaluate the ability of the entire system (hardware, software and human components) to meet the performance criteria that were assumed in the safety case.

CBTM Example

All the sources of uncertainty discussed above regarding how the ultimate system would impact human performance arose in the CBTM case study. The HRA took place when the CBTM implementation was in an early prototype stage, with limited field-testing. The local train crews who participated in the human factors quantification workshop had more experience with CBTM than anyone else, but their experience was still limited: they only operated a CBTM equipped train a couple of times each. Further, the CBTM was still in the early development phase and would be expected to improve substantially prior to actual implementation. As a result, experience with the prototype version of CBTM was not necessarily representative of performance of the final production system. In addition, operational details related to training, procedures and policy were still to be worked out.

All these unknowns made it difficult to establish definitive human error estimates. While a number of different opinions were expressed and evidence offered, there was a general consensus that there was not sufficient experience with the CBTM system to make confident projections of its likely impact.

Two strategies for coping with sources of uncertainty regarding the likely impact of final system implementation on human performance are: (a) make specific assumptions about the system that will ultimately be implemented, and generate human reliability estimates based on those specific assumptions; and (b) perform sensitivity studies to explore the safety consequences of alternative assumptions about the final system implementation and its impact on human performance. Both these strategies were adopted in the CBTM case study.

The quantification workshop participants noted a number of human factors deficiencies with the current CBTM prototype, including inaudible alarms, a high false positive alarm rate, and lack of consideration of dynamic braking. For human reliability quantification purposes, the workshop participants explicitly assumed that these nuisance elements of the prototype system would be eliminated in the production version of CBTM.

At the same time, an explicit recommendation of the quantification workshop was to perform sensitivity analyses to explore the safety consequences of alternative assumptions regarding the impact of CBTM on human performance. The human reliability estimates generated at the workshop for CBTM were recommended as starting points for the sensitivity analyses.

Ultimately the actual impact of the final CBTM implementation on human performance should be assessed in person-in-the-loop tests to validate the assumptions made in the safety analysis. For example, a design assumption of CBTM is that train crews will be able to operate the trains in such a way that CBTM automatic braking will rarely be initiated. However, the braking algorithms in the current CBTM prototype require that locomotive engineers brake earlier than they normally now do. As a consequence, CBTM warnings and penalty braking are often activated. It remains an empirical question whether: (1) The braking algorithms will be improved in the final implementation of CBTM so that CBTM will no longer require locomotive engineers to brake as early as they must now; and/or (2) locomotive engineers will be successfully trained to change their braking strategies so that they conform to the braking curves assumed by CBTM.

4.6 Expertise Needed for a Human Factors Quantification Team

One of the important elements to the success of a human factors quantification project is to assemble an interdisciplinary team to conduct the quantification that jointly possess experience and expertise in:

- Human Factors
- Human Reliability Analysis
- Probabilistic Risk Assessment
- Group Facilitation Techniques

It is important to include human factors expertise on the team to support the conduct of the qualitative analyses that feed into the human performance quantification. Relevant human factors expertise includes knowledge of:

1. Interview and observation methods;
2. Factors that can contribute to human performance and human error; and
3. How new technologies can both positively and negatively impact human performance (based on lessons learned in the railroad and related industries).

Another complementary set of skills required on the team is knowledge and experience in the performance of human reliability analyses as part of probabilistic risk assessment programs. It is important that one or more members of the team be fully versed in techniques for generating human reliability probabilities, the pitfalls to watch out for (e.g., assuming that events are independent when in fact they are not), and the commonly accepted standards and practices in the human reliability analysis field. The results of the human reliability analyses are often used as the basis for important regulatory decisions. It is important that the human reliability analysis methodology employed meets the standards in the field and can withstand the scrutiny of peer evaluation.

It is also important that the team includes one or more members that understand how human reliability analyses fit into the larger probabilistic risk assessment process. The rationale of conducting a quantitative human reliability analysis is to support a larger scope quantitative risk assessment. There needs to be someone on the HRA team that can ‘speak the language’ of the

PRA community and serve as a bridge between the human factors and human reliability analysts that concern themselves with human performance and the PRA specialists that are attempting to integrate human and equipment reliability estimates to form overall risk estimates.

Finally, it is important to include on the team one or more people that are skilled at group facilitation and methods for eliciting subjective probabilities from human experts. As reviewed in Section 2.3.5, there is an extensive body of knowledge for eliciting subjective probabilities from human experts. This body of knowledge is based on a combination of scientific research on human judgment and biases as well as pragmatic experience in developing and applying methods for risk assessment in the decision sciences literature. It is important that one or more members of the human factors quantification team be familiar with the kinds of biases that can arise when eliciting probabilities from human experts and the methods available for minimizing those biases.

While the skill set called out includes human factors, human reliability assessment, probabilistic risk assessment, and group facilitation skills, it does not mean that the human factors quantification team requires four different people, each with a distinct set of expertise. It may be that a given team member will have expertise in more than one area. What is important is that this set of knowledge and skills are well represented in the team.

CBTM Example:

In the CBTM case, the human factors quantification team possessed extensive experience and expertise in human factors, human reliability, probabilistic risk assessment, and group facilitation.

The field observations and interviews were led by a human factors specialist with expertise in the conduct of field observation studies and focus groups. The team conducting the field observations and interviews also included a second human factors expert with expertise in the impact of technology on human performance as well as a human reliability analysis expert.

A human reliability analysis expert with extensive knowledge and experience in reviewing human performance and incident report databases led the human performance database review. An expert in PRA supported the HRA expert.

The human reliability analysis expert led the human factors quantification workshop with active support from the two human factors experts and the PRA expert, who also had extensive expertise in subjective probability estimation methods and group facilitation techniques. Because of the large number of participants in the Quantification Workshop, another human factors specialist who had extensive experience and skill in-group facilitation processes augmented the team.

The human reliability and probabilistic risk assessment experts computed the human reliability values jointly. Their expertise in human reliability quantification methods and tools and risk quantification requirements, enabled them to prepare results in a form that could be integrated into the probabilistic risk assessment process. Finally, this report was generated as a joint effort of all the team members.

4.7 Integrate Results into PRA

Our task was to perform an HRA that would be compatible with the larger CBTM probabilistic risk assessment. The issue of compatibility is a key one. The PRA must be structured in such a way that it asks questions that can be answered by HRA; the HRA must provide quantitative results that are appropriate to the context of the PRA.

In previous sections, our report has focused on the needs, capabilities, and approaches to HRA. This section focuses on the integration of the results of the HRA into the broader PRA. Appendix D provides a brief description of PRA, how it is used, the various approaches, and their advantages and disadvantages. Here we discuss how the HRA ought to be integrated with the PRA, describe the challenges that were faced in the CBTM case, and offer alternatives for how the HRA results reported here can be integrated with the ASCAP PRA simulation model.

The philosophical structure of PRA is consistent with the approach to uncertainty taken in the CBTM HRA. This consistent framework ensures that the results of the HRA are compatible with the general quantification structure of a PRA. The events quantified in the HRA conducted for CBTM correspond to 'human failure events.' These human failure events are basic events in the PRA model (i.e., events analogous to equipment functional failures in the way they fit in the logic model of the PRA and the way they combine with other events in the quantification of the PRA model).

As discussed in Section 2.2, human failure events are combinations of specific unsafe actions each occurring under specific contextual conditions. At the level of the unsafe actions, the HRA must develop the quantification consistent with the context, and in light of the cognitive processes affecting human performance (Reason, 1990; 1997). The qualitative analysis that occurred prior to and during the quantification workshop served to bring out the range of unsafe actions that can occur and the contexts in which they occur. The quantification process built on this analysis to generate probabilities for the higher-level human failure events that aggregate across unsafe actions and contexts. The probabilities for these human failure events serve as the input to the PRA.

CBTM Example
In the case of the ASCAP model for CBTM, there are some specific interface challenges between the output of the HRA and the ASCAP model. The PRA simulation model was constructed before the HRA team was assembled. The level of decomposition of the human failure events in the simulation model is more detailed than the level at which the HRA could be meaningfully quantified (See Section 5.1). Two alternatives for integrating the results of the HRA into the PRA quantification could be adopted. The simplest, from an overall model standpoint and for clarity for review, would be to slightly restructure the PRA simulation model such that the human failure events in ASCAP match the events quantified in the HRA.

If the effort required to restructure the PRA simulation model is extensive and costly, there are work-arounds to permit quantification. One suggested work-around is to use the human failure event probability as the input for one of the lower level elements within ASCAP[13]. Other parts of

[13] The ASCAP human factors model combines probabilities of multiple lower level elements (e.g., probability that train crew is responsive, covered, compliant, that dispatcher is responsive, covered, compliant) to obtain the probability of a human error event (e.g., train fails to stop at boundary of authority). The suggested 'work-around' is to assign the probability of the higher level human failure event produced by the present HRA to one of the lower

the ASCAP decomposition model could be short-circuited by assigning probabilities of 0.0 or 1.0 as needed to have the same net result as rebuilding the ASCAP model.

The second work-around involves more work and is less transparent. The HRA human failure events could be decomposed to the level currently in ASCAP and pushed through ASCAP quantification. In this workaround, probability values would be created for each of the lower level elements included in the ASCAP model (e.g., probability of an agent being responsive, probability of coverage, probability of an agent being compliant) so that when the ASCAP simulation model was run it would generate the results of the present HRA for the higher level human error events (e.g., probability that train fails to stop at boundary of authority).

This work-around has a number of serious drawbacks, and is therefore not recommended. First, it may take several iterations to reach the point where the calculated ASCAP results match the original HRA results. Second, there is no guarantee that the probabilities assigned to the lower level elements have validity. The probabilities for the higher level human failure events that were produced in the present HRA have some degree of validity, in that they have been derived by a systematic method that combined values obtained from operating experience databases and expert judgment of individuals with relevant operating experience. There is no guarantee that the probabilities assigned to the lower level elements in the ASCAP model would have the same degree of validity. If the only objective of the exercise is to replicate the results of the HRA, there is no problem. However there would be no basis for utilizing these probabilities in a different context or different analysis.

Lastly, a problem with attempts to decompose higher level human failure events into lower level elements, is that important dependencies among the lower level elements may be missed (see Section 5.1). For example factors that might increase the probability of a communication error (e.g., a noisy radio channel) may also decrease the probability that that error will be caught and corrected (i.e., the probability of coverage). Thus, it may be inappropriate to treat these probabilities as independent.

level elements and then assign ‘0.0’ or ‘1.0’ to the other lower level elements so that when the probabilities are combined within the ASCAP human factors model the resulting value for the higher level human failure event matches the probability value generated by the HRA. The intent of this ‘work-around’ is to insure that the appropriate probabilities are generated by the ASCAP model for the higher level human failure events, without requiring extensive changes to the ASCAP model and software.

5. RECOMMENDATIONS & CONCLUSIONS

5.1 Findings from CBTM Analysis

5.1.1 General Findings.

The approach taken for this HRA generated reasonable results despite the fact that there was no directly applicable database. The workshop format permitted experts from many different organizations and backgrounds to work together and reach consensus. Uncertainty was expressed through probability distributions that were accepted by the group. The HRA and PRA/ASCAP teams reached agreement that the HRA results could be incorporated in the ASCAP model (i.e., they are appropriate for use in the PRA).

5.1.2 Potential Limitations and Concerns

Although the approach worked well, there were several areas of concern that need to be pointed out. These include biases in data, the level of modeling of human error events, and modeling of future CBTM operations. These concerns do not limit the value of the HRA; they do, however, require that analysts be alert to their effects, and adapt the analysis as required to account for these potential problems.

Biases in Data

Any uses of data, either from databases of operational experience or from the opinions of experts, have the potential to contain biases that lead to incorrect estimates of probabilities. Databases, as discussed in Section 2.3.4, are subject to various limitations because in almost all cases, the data have been gathered for reasons other than supporting a PRA or HRA study. Because of these differences, events of concern to the PRA and HRA may be missing, underreported, or over reported. For example, the events in the FRA incident database must involve certain criteria associated with the consequences of the incidents. However, our analysis is concerned with the causes of incidents, many of which will not involve the consequences for the events to be reported to FRA. Similar opportunities for gaps or misalignments exist for the other databases available for this project, such as the disciplinary databases.

Our approach has been to review these databases for potential limitations and biases in the reporting requirements for the databases, review these with the workshop attendees, and making filtering and scaling adjustments based on the inputs of the participants. However, we recognize that these adjustments represent opinions and the adjusted values may still contain biases. As discussed below, we took steps to limit the potential for significant biases in these opinions, but there is no guarantee that the results are entirely free from bias.

Experts are subject to limitations in their ability to consider and use all the data in their experience, as discussed in Section 2.3.5. They may focus on more recent or other limited sets of experience, and ignore the experience of events that occur very rarely and have not been experienced recently. The use of databases (despite their own limitations outlined earlier) act to remind people of experience from elsewhere that may have not been seen personally by the people assisting in the workshop, but which nonetheless could happen. In addition, we used elicitation processes that are intended to limit to the extent possible these kinds of biases. We believe that the results are sufficiently accurate for the purposes to which FRA will use the PRA and HRA results, particularly since the values explicitly include uncertainty ranges that the participating experts felt encompassed the reasonable state of knowledge.

Level of Modeling of Human Error Events

The modeling in the human reliability analysis task was different from that embedded in the ASCAP model. The ASCAP modeling decomposed human error events to a smaller level of decomposition, explicitly modeling errors in perception and action, and failures to recover ('coverage') from these errors. The modeling in the HRA task simply estimated the likelihood of the outcomes of concern, such as entering a block for which the train has no authority. The reasons for this level of modeling in the HRA study are described in Section 4.1. As a result, the potential exists for a 'mismatch' between the ASCAP and HRA modeling. Recommendations for ways to deal with the potential mismatch are provided in Section 4.7.

Modeling of Future CBTM Operations

Part of this work included estimating the likelihood of various errors while operating the CBTM system. Our ability to predict the likelihood of errors with confidence was limited by a variety of factors. These factors include the following items:

1. CBTM system was still undergoing field trials,
2. Its design was not finalized,
3. Only a limited number of engineers, conductors, and dispatchers had experience with the system.

To overcome these limitations, we held lengthy interviews with engineers and conductors who had experienced the trials of the CBTM system. They identified potential areas of design and operation that might result in errors or other operational problems. These interviews were based on our knowledge of the human-performance concerns that can occur when a new technology is introduced into an existing operating environment, as discussed in Section 3.3. These interviews aided the discussions among the workshop participants in considering the range of potential errors and suggesting bounds on their effects. These interviews and workshop did not guarantee that the results are precise (as discussed in Section 4.5), but we believe that the potential issues for future operational concern were identified and discussed as effectively as possible.

5.2 Recommendations for Future HRA Studies

The analytical situation that arose in the present study, having relevant data but with a variety of limitations (i.e., data sources that may lead to both underestimates and overestimates of frequency) are far from unique. Limitations in the data occur often in the railroad industry and other industries, and must be addressed explicitly.

The approach we took for combining 'hard data' with expert judgment is a good approach that could be used in other applications. It uses 'hard data' to ground the experts judgments, while using expert judgment to compensate for the known limitations of the existing data.

Guidelines for human factors and human reliability analyses were developed as part of this project). The guidelines are intended for organizations developing an HRA plan as well as regulatory agencies such as the FRA charged with evaluating an HRA analysis submitted as part of a product safety plan. The primary recommendations are summarized here and described in more detail in Appendix E.

1. Use an HRA team that includes members experienced in performing human factors studies, human reliability analyses, probabilistic risk assessments, and group facilitation.
2. Model human errors at compatible levels in the PRA and HRA tasks, preferably at the level of available data and experience.
3. Verify that the data sources (databases, expert judgment or a combination) are suitable for the tasks and associated errors being analyzed. Identify gaps or mismatches and utilize expert judgment to leverage the available data while compensating for the known limitations of the data.
4. Conduct qualitative task analyses with people experienced in using the existing systems. Activities should include interviews with workers using the existing systems or the target users of the system (in the case of technologies under development), their trainers and supervisors, so that all levels of experience are included.
5. Expert elicitation methods should take into account known biases and other limitations of expert judgment. Experts should express their opinions in terms of ranges rather than single point values.
6. Solicit input from as broad a range of stakeholders as possible so that the analysis takes into account a wide range of perspectives. Accept quantitative inputs only during the elicitation process, from people with relevant operating experience.
7. Ask the broadest range of stakeholders possible to review to the *results* of the analyses to foster support for the results.

5.3 Conclusions

The approach taken in this study provides one viable approach for others to perform HRA studies in support of the FRA's proposed rule: Standard for Development and Use of Processor-Based Signal and Train Control Systems. The lessons learned from this analysis of the CBTM system have been documented and can provide the FRA or others interested in performing or using human reliability analyses with guidance on avoiding potential pitfalls in future studies.

APPENDIX A. SUMMARY OF LOCOMOTIVE ENGINEER AND CONDUCTOR INTERVIEWS

Introduction

This appendix summarizes the results of interviews and observations of CSXT locomotive engineers and conductors conducted in Spartanburg, South Carolina, on April 18 and 19, 2000. The interviews and observations were conducted as part of a Human Factors study in support of human reliability modeling as input to ASCAP. The objective was to get a deeper understanding of the complexities that affect locomotive engineer performance, potential for error, and how CBTM is likely to affect locomotive engineer performance and impact safety.[14]

Activities included:

- Observations during a 4 hour head-end ride on April 19 that were conducted as part of the CBTM test;
- Two-hour interview with a locomotive engineer that had served as a CBTM trainer, introducing locomotive engineers to the CBTM system.
- Two two-hour focus groups of locomotive engineers and conductors. Eight individuals (6 locomotive engineers and two conductors) participated in the focus groups.

The engineers and conductors participating in the focus groups ranged in level of experience from 28 years to 11 months. They also ranged in experience with CBTM from operating trains with CBTM installed on numerous occasions spanning the period it has been piloted, to having been on only one train run with CBTM.

The focus groups were conducted in an off-site conference room and the locomotive engineers and conductors participated voluntarily on their own time.

The interviews/focus group sessions addressed two main topics:

- Factors that make running a train challenging in today's environment and potential for error.
- Impact of CBTM on train crew performance.

The remainder of this appendix is divided into two sections.

[14] Failures to take action to prevent hazardous events, and actions that cause hazardous events, are commonly called "human errors" in quantitative risk assessments. This term does not imply that people are necessarily personally responsible or culpable in some way, just that an action was omitted (or taken) that had an adverse influence on safety. We realize that these so-called 'errors' are often the result of particular circumstances or conditions in the workplace. They may be work-related (such as fatigue at the end of a long shift or the result of unusually high workload) or environment-related (such as a temporary sign being obscured by heavy rain or fog). They may also be caused by equipment problems (such as mishearing dispatcher instructions over poor radio links in certain locations) or organizational factors (e.g., policies, work rules). The purpose of the interviews and observations was to begin to understand and document the kinds of errors can plausibly occur and the range of factors that contribute to them.

Section 1 provides an overview of the most important findings, and the potential implications for modeling the impact of CBTM on human performance in ASCAP.

Section 2 consists of detailed notes that integrate the feedback obtained from locomotive engineers and conductors across the interviews and focus groups. It also provides the detailed evidence in support of the quantification

1. Main Findings: CBTM Impact on Train Crew Performance and Safety

- All eight individuals interviewed in the two focus groups and the trainer that was interviewed (a total of nine interviewees) felt that CBTM has the potential to improve safety and would like to see it implemented.
- They believed that CBTM would provide effective support in cases where they might forget to reduce speed or stop at end of authority due to attention lapses or failures of memory.
- They particularly liked that CBTM warns them when they are about to enter a work zone and when temporary speed restrictions are in effect. These are cases where the probability of error is likely to be higher and the consequences of an error may be severe.
- All nine individuals interviewed also perceived some limitations of the current pilot version of CBTM that need to be addressed before its benefits could be fully realized.
- Opportunities for improvement fell into three classes:
    1. The audio alert was easy to miss given the noisy cab environment (e.g., engine noise, the whistle, the radio, conversations) and the CBTM display is outside of the primary field of view. This had two consequences:

       Sometimes they missed the audio alert, resulting in a penalty brake application that could have been avoided had they noticed that an information message (requiring a yes/no response) or warning had come on.

       Because of the severity of consequences of missing an information message or warning (i.e., the penalty brake application) the locomotive engineers felt a need to continuously monitor the CBTM display – which is an added attention burden that they felt they could not afford and could potentially distract them from attending to events outside the cab (e.g., trespassers, motor vehicles at grade crossings).

       This issue can be easily addressed, for example by selecting an audio alert that is easier to detect and discriminate from other sounds in the Locomotive Cab. One solution is for the audio alert to stay on until the locomotive engineer explicitly acknowledges it.
    2. In many cases, the warning message does not come on early enough before the penalty brake is applied to allow the locomotive engineer to respond in time to avoid the penalty brake. This reduces the ability of the locomotive engineer to take advantage of the warning message. It also reduces the ability of the human agents (locomotive engineer and conductor) to catch and recover

from any ‘errors’ that the CBTM system might make. Thus, it reduces their potential to serve as a recovery mechanism.

More time is needed between the on-set of the warning and the initiation of the penalty brake to allow the locomotive engineer time to slow down the train to the appropriate speed and/or select an appropriate stopping place.

3. Often the CBTM system determines that braking is required at an earlier point than the locomotive engineers would themselves choose to start to brake. In addition, in some cases, a warning comes on in situations where the locomotive engineers felt stopping was not necessary or appropriate. In some cases, the position where the CBTM stopped the train was inconvenient, making it hard to restart the train. Stopping at an inappropriate time or place may also introduce a new source of risk.

- With respect to the issue of potential for complacency and over-reliance on CBTM, the engineers provided mixed reactions. On the one hand, they indicated that it remained their responsibility to make sure that no movement authorities or speed restrictions are violated, independent of whether they were reminded by CBTM or not. The analogy one locomotive engineer gave was to an advanced warning board on the side of the track. If it is there, it can help the locomotive engineer remember that he will need to brake soon. However, if for some reason it is not there, the engineer is still responsible for remembering to brake. The same would be true for the CBTM system. It would provide an aid, but the engineer would know that the ultimate responsibility remains on his/her shoulder. At the same time, the engineers noted that if the CBTM system was there and was working well they would tend to rely on it. As one engineer put it “If we can’t rely on it, I don’t want it up there. If it works I’ll rely on it.”

- With respect to whether CBTM would change the behavior of the locomotive engineers, the locomotive engineers indicated that it would. Given that the CBTM system expects the locomotive engineers to start to brake earlier than they are now inclined to, they would need to learn new braking styles. This raises a need for training not only on the CBTM interface and how to use it, but also training on train handling and braking that is more consistent with the expectations of CBTM.

- locomotive engineers reported that the interface for entering consist information into CBTM was easy to use. When asked whether locomotive engineers might intentionally enter incorrect consist information in order to manipulate when the CBTM system came on, all 9 individuals interviewed felt that that was very unlikely. They indicated first that CBTM contributed to safety and they wouldn’t want to take action to defeat that, and second, since it is a computer system, it records all inputs, and it would therefore be easy to catch when someone did this.

- Because the CBTM display is only on the locomotive engineer’s side, the conductor cannot see it. However, in the current task allocation between locomotive engineer and conductor, the conductor has responsibility for serving as a redundant check/reminder to the engineer. Several of the individuals interviewed argued that it would be helpful to have CBTM displays on the conductor's side as well as the side of the locomotive engineer.

These are the main findings that relate to how CBTM is likely to impact the performance of locomotive engineers and conductors. There were additional detailed findings on the types of human errors that arise, the factors that contribute to them, and how a system such as CBTM could help catch and recover from those errors. These results are presented in the detailed integrated notes in the next section.

Implications for CBTM and Modeling of Impact of CBTM on Crew Performance in ASCAP

Several of the issues raised by the locomotive engineers with respect to the pilot version of CBTM (e.g., the audio alert is difficult to detect; there isn't sufficient time between when the warning message comes up and when the penalty brake is initiated) can be easily addressed so that they are no longer an issue when the final CBTM system is implemented.

The CBTM system clearly includes provisions for catching and recovering from human errors. However, the interviews provided some suggestion that CBTM might introduce new sources of risk by stopping at an inappropriate location. If the audio alerts are improved, and the length of time between the warning message and the time the penalty brake is activated is lengthened to allow the train crew to take action to prevent a penalty brake, then the human crew can also play a role in catching and recovering from 'errors' made by CBTM.

With appropriate warning, the train crew can take action:

- To avoid stopping when it is not necessary (e.g., when the locomotive engineer has authority to move into a territory)
- To choose WHERE to stop to minimize the potential for delays, inconvenience and safety vulnerabilities.

The ASCAP team may want to consider modeling this source of 'coverage' as well.

The current ASCAP model assumes the 'simplest case' version of communication between dispatchers and train crews. In particular, it assumes that the train crew initiates requests for block authority and that dispatchers issue authority for a single block at a time. However, interviews and observations indicated that:

- Sometimes it is the dispatcher who calls the locomotive engineer to provide authority [without the engineer calling him/her first]
- Sometimes the dispatcher will give authority for multiple blocks simultaneously
- Sometimes the dispatcher will tell the train crew not to call the dispatcher to release blocks after passing them one at a time but to wait for the dispatcher to call asking what blocks they have passed.

These dispatcher behaviors have to do with the fact that the dispatcher has a heavy mental workload. As a result he/she tries to be proactive, to perform tasks during lower workload periods.

Future versions of ASCAP may want to model these more complex dispatcher – train crew interactions.

These findings also suggest that it may be valuable to eventually model dispatcher workload in ASCAP, where workload is a function of number of trains he/she is currently handling, and the likelihood of dispatcher error is a function of workload.

2. Detailed Notes Based on Interviews, Focus Groups, and ‘Head End’ Ride Observations

Background:

- Approx. 70 engineers have received introductory training on the CBTM system. This training consists of observation of a prototype CBTM in an office setting and demonstration of the different user interfaces that come up under different simulated conditions. Training did not include experience running a train that was equipped with CBTM.
- Currently six locomotives are equipped with CBTM. These do not necessarily stay in CBTM territory however. As a result, individual locomotive engineers have not necessarily had much experience with operating a train equipped with CBTM.

Characteristics of CBTM:

- Audio alerts:
  - One beep indicates an informational display, a prompt for crew response with no braking imminent, or a warning with no braking imminent.
  - Two beeps indicate that braking will start soon if no corrective action is taken.
  - Three beeps indicate that braking has started.

Prompt for crew response with no braking imminent occur in cases where the CBTM does not have the information to know whether the train has authority to proceed. An example is: ‘Do you have authority to enter this work zone?’ This prompt occurs because the CBTM may not have the information about whether the locomotive engineer has obtained permission from the roadway worker in charge to enter the work zone. This prompt appears approximately 3 miles away. The locomotive engineer must respond yes or no.

Currently if the ‘CBTM’ is not operating properly (giving too many ‘nuisance’ alerts) then the locomotive engineer can isolate it. In fact, some of the locomotive engineers said that when they had difficulties and called the dispatcher, the dispatcher told them to isolate it. When the CBTM system is isolated, it still displays messages, but does not provide audio alerts and does not apply the penalty brake.

Feedback from Locomotive Engineers and Conductors on CBTM:

- All eight individuals interviewed in the two focus groups and the trainer that was interviewed felt that CBTM has the potential to improve safety and would like to see it implemented.
- However, all nine individuals also felt that there were problems with the current pilot version of CBTM that needed to be addressed before its benefits could be fully realized.

The following summarizes locomotive engineer and conductor comments.

Audio Alert Easy to Miss/ Perceived Need for Constant Monitoring of the CBTM Display:

- The audio cue is easy to miss particularly because there is lots of noise in the Locomotive Cab (e.g., the whistle, the radio, conversations).
- The display is above your head and you can't easily see it (from peripheral vision)
- You are busy handling the train and can't always be looking at the box (CBTM interface).
- I think it would change how you behave. We are thinking much more about that machine. Have an additional computer screen to look at. Could cause you to lose your train of thought.

The time between warning and activation of the penalty brake is sometimes too short:

- If running at 40 miles an hour and CBTM wants you to be at 10 miles an hour then the time between the warning and the automating braking is not enough time (especially a fully loaded train – braking speed depends on tonnage, train length and gradient).
    - First service brake on a fully loaded train has minimal effect o n the speed and can't bypass the first service brake.
- You can start to use brakes to slow down but if the CBTM does not see the speed change, it will stop the train.
- One engineer suggested looking at how warning signals are implemented in Traditional train control signals. He indicated that in those systems the warning signal comes earlier relative to when the penalty brake is initiated (He mentioned sixty seconds time interval).
- When the warning comes in, there is not enough time to slow the train down enough to avoid the penalty brake. Even if put full brake application; there is simply not enough time to slow the train down.

CBTM sometimes comes on too early or when the locomotive engineer does have authority to proceed:

- Some places it is making us do things "that is not the way I would do it personally"; "will make you do something earlier than you normally do it". Examples:

The engineer may be going 40 miles an hour and know he/she needs to get to 25 miles in three miles. He/she may not feel the need to start slowing down yet. CBTM however, might come on with a warning followed by a penalty brake. Once the warning comes on, often the engineer is not able to slow the train down quickly enough to prevent the penalty brake from coming on.

At end of a block (end of block authority), it wants you to slow down sooner than the engineer might be inclined to start slowing down.

It stops you too early. It doesn't pick up the dynamic brakes

Sometimes CBTM indicates that you are about to enter a block without authority, when in fact have gotten movement authority from the dispatcher for several blocks, and therefore have authority to move into the next block. The locomotive engineers indicated that dispatchers routinely give multiple blocks but that CBTM seemed to not know that.

Problems associated with stopping too soon/at wrong place:

- If stop too soon may be blocking an intersection and other trains behind can't move
- Some places if you stop, it might be a real challenge to get started again (e.g., on a hill, if stop before crest the hill, may have to back up)
- After stop, may not have full braking capability after immediately restart – if you have to stop very quickly soon after restart it may be a problem.
- You can't start and stop anywhere. You can stall.
- Sometimes you are coming on an uphill with a heavy train (14,000 tons). You are 2 to 3 miles from where you need to stop. The CBTM system tells you that you need to slow down – but in fact, you need to pick up speed (and not slow down) in order to make it up the hill (at which point you can slow down and still stop in time.)
- If you are coming down a hill and the CBTM stops you too soon, you may not have enough brakes (air all gone) to stop again at the bottom (where you need to be stopped).
- If it is a heavy train in the rain, it could be dangerous to stop at a hill (it could derail, it could buckle).

Factors that could lead the CBTM to apply penalty brake erroneously:

- Braking strategies that are too conservative (or take limited set of factors into account)

In cold weather (winter) a train doesn't stop as fast as in warm weather (summer). Does the CBTM algorithm take this factor into account? If it doesn't then if you design for winter weather you will brake too early in summer. If you design for summer weather you will not brake soon enough for winter conditions.

- Inaccurate information:

Dispatcher might put in the wrong information into the system:

  - A case was described where a dispatcher entered in the wrong Engine number into the computer. He gave the correct train verbal authorization to proceed, but entered the wrong Engine number in the computer. This happened around a shift change and was later caught by the incoming dispatcher when he questioned the train (said to them you don't have the authority to proceed – and they said that they did.)

locomotive engineer might fail to update consist information in CBTM when the crew adds or removes cars/cargo.

- CBTM warns and enforces the non-violating train when it determines that an equipped locomotive is occupying a block for which it does not have authority (Violated Authority Warning). A penalty brake applied on the non-violating train might cause the non-violating train to stop at an inappropriate place (if the locomotive engineer fails to detect the warning and the penalty braking is applied.)
- Failures of locomotive engineer to notice and answer informational messages:

If the locomotive engineer fails to detect the informational message (that is signaled by a single audio beep, which is easy to miss given the noisy environment) then he/she will not respond, this will cause the CBTM to display a warning which can also be missed (especially since the locomotive engineer knows that he/she has authority to proceed and may not be expecting this warning) as a result a penalty brake will be applied when the train had not violated its authority.

Potential for Complacency: Two perspectives were expressed:

- "If we can't rely on it, I don't want it up there. If it works I'll rely on it."
- It plays the same role as advanced warning board, if it is there it is a help – but the locomotive engineer and conductor remain responsible for abiding by the speed restrictions and movement authority whether or not the advanced warning board is there – CBTM is similar. If it is there it is an aid, but the train crew will continue to exhibit vigilance.

Memory aids such as 'advance warning boards' and 'written train bulletins' offer a useful parallel in thinking about the potential for complacency. If CBTM is working well and the warnings come on appropriately, then they will serve a similar role to an 'advance warning board' – they will provide a reminder, reducing the potential for errors due to memory lapses. If the CBTM system is not functioning, and there are clear visual cues that it is not operational, then the locomotive engineer will know that he cannot rely on it, and the probability of error reverts to the probability of error in the base case (may want to argue that it is slightly higher because of loss of skill). However, should the CBTM appear to be functioning, but for whatever reason does not provide a warning (reminder) that a speed restriction/end of authority is coming up, then there is likely to be some effect of 'complacency' (i.e., reliance on the reminder) and the probability of human error is likely to be somewhat higher than the base case.

A potential issue of complacency arises when a train that is equipped with CBTM passes through territory that is not connected to CBTM. The locomotive engineers mentioned that there are several miles in the Monroe Sub (3 to 6 miles?) that occur in the middle of the CBTM territory that is not connected to CBTM. In those cases, if the CBTM display does not make perfectly clear that it is not functioning in that territory, there may be a potential for increased human error associated with 'complacency'. In the words of one locomotive engineer, "If it is not 100% effective it won't be helpful."

Specific Comments of locomotive engineers on whether CBTM affects how they operate the train:

- Yes – Have to do things earlier to satisfy the system. Need to do some things differently than you would normally do them to avoid the system coming on and applying a penalty brake.

- Right now CBTM doesn't recognize dynamic braking.
- Yes – CBTM would affect train handling
- Yes – "We are thinking more about that machine (CBTM), have a computer screen to look at, and could cause you to lose your train of thought."
- Yes – A new person (less experienced engineer) could learn this new style of train handling (slowing down sooner) but for older people, as the saying goes 'you can't teach an old dog new tricks".

CBTM Interface for Entering Consist Information:

- Simple user interface – easy to enter consist information, and required information is readily available
- When the consist changes (pick up or drop off cars or cargo) will need to update the information in CBTM:

CBTM does provide a reminder to do this in many cases –

However there may be a possibility of forgetting to update it – possibly leading to braking too late [or possibly as bad – too soon]

- The trainer indicated that on the Spartanburg to Augusta territory this is not so much of an issue but in other territories, locomotive engineers might be changing cars in and out four or five times.
- On the question of whether locomotive engineers would be inclined to intentionally put in incorrect consist information – the answer is not high probability:

Could happen to try to delay when CBTM decides that braking is needed but:

  - Not likely to be an issue if you are told not to tamper with a safety device
  - Once you tamper with a computer device they can track you down (it records the Locomotive inputs – so you can be easily caught)

- However there is still potential for errors in the consist information in the CBTM due to data entry errors or forgetting to update it when cars are taken off or added.

Other concerns expressed regarding CBTM:

- CBTM could be setting the stage for a one man crew or even 'remote control' where you don't have an engineer at all

Summary comments on the CBTM system performance and opportunities to enhance it:

- System has problems that need to be improved because it can be a nuisance and a hazard right now.
- CBTM is a good system. It is going to be good and safe. However, it needs improvements to the audio alerts and interval between when the warning comes on and when the penalty brake is initiated:
- Engineers and conductors are busy and need an earlier alert (time between when the warning comes on and the penalty brake)

- Yes CBTM is good for safety. Would like to see improvements with regard to the terrain information it includes and uses in braking – so that brake won't be applied as quickly if it is on flat ground.
- Overall it will be a good system if the quirks are worked out and it is not viewed as a replacement but an enhanced of the other co-workers (Roth Note – this refers to the conductor)
- Anything that helps the safety of movement of trains is good. CBTM is on the right track but there are still some 'kinks' in it.
- Right now, it results in increased workload for the engineer in having to monitor and interact with the CBTM interface. Because the CBTM is on the engineer's side, the conductor cannot help. Consider putting a CBTM screen on the conductor's side too.
- Good idea to have a system set up to protect train collisions as long as the system is reliable.

Factors that contribute to errors such as exceeding speed limits or going past Limits of Authority:

- Attention Lapses/Distraction/'Mental Vacation':
    - Had a head-on collision 10 years ago because the locomotive engineer and conductor were talking and went past the block of authority.
    - Another case (in signal territory), passenger train missed an approach (it was really foggy; they weren't expecting an approach signal) and their attention had been diverted by a school bus that was coming toward a grade crossing
    - Once a locomotive engineer exceeded block of authority (started to head back without getting authority) just due to an attention lapse. This error was caught by the dispatcher who overheard on the radio that they had started to head back and alerted them that they did not have authority
- Memory lapse (forgetting): Particularly vulnerable in cases where the speed restriction or stop location is temporary. [When it is permanent, then the locomotive engineer is likely to remember that it is there since due to training and repeated operation over the same territory the locomotive engineer is likely to have a good mental model of the territory and where the permanent blocks and speed restrictions are located.]
- Slow Orders: Temporary speed restrictions:

Especially if recently issued

Especially if issued after the train bulletin that lists temporary speed restrictions is issued so that the engineer is provided the information verbally over the radio by the dispatcher

Especially in cases where the speed restriction signs have not yet been put up [or are obscured]

Especially when the time duration between when the dispatcher provides the information and when it comes into effect is long (e.g., will come to the speed restrictions four hours after the dispatcher called to tell him/her about the speed restriction)

- Work Orders: If roadway workers "own" a territory a by (707 Authority), then the locomotive engineer needs permission from the roadway worker in charge (via radio communication) prior to entering that territory. – However, the engineer can forget to request permission from the roadway worker in charge.

Factors that contribute to the likelihood of forgetting are the same as for slow orders.

- Confusion about where the speed restriction or stop location is. Particularly vulnerable in cases where the speed restriction or stop location is temporary.

Particularly vulnerable when the location is between mileposts – so that there may not be visual cues as to where the location is.

Particularly vulnerable when visibility is poor (night, poor weather) so that visual cues to aid in identification of location are degraded.

In the case of permanent speed restrictions and block end of authority, then the locomotive engineer is likely to know exactly where it is due to training and repeated operation over the same territory. The locomotive engineer is likely to have a good mental model of the territory and where the permanent blocks and speed restrictions are located. He/she is likely to have multiple cues to help identify the location (not only mileposts, but also 'land marks' such as Houses and trees, and non-visual cues such as vibration, curves, inclines)

- Communication errors:

Errors in communication do happen but there is typically ample opportunity to detect and correct them.

When riding in the Locomotive Cab for the CBTM test, an error in communication that was caught occurred: The conductor gave the wrong time (said 12:35 when in fact it was 13:35) to the dispatcher when he called to release a block. The dispatcher didn't immediately catch the error, but a road foreman who happened to be on the train and was listening to the conversation caught it and corrected it and the right information was communicated to and read back by the dispatcher. This incident illustrates that communication errors likely occur with relatively high frequency but that in the large majority of cases they are caught and corrected quickly.

- Improper train handling – know where you are supposed to stop or what speed you are supposed to go at but overshoot due to braking too late or insufficiently. One example given was start to put the brake on too late, slack coming in will push you forward (past where you intend to stop)

Factors that can increase the potential for human error:

Impact of Fatigue:

- Can be slower to react (detect, put on the brakes, etc.)

Impact of stress:

- Serves as a source of distraction

Impact of level of experience:

- Braking skill:
- Confidence on stopping distance at any given speed is lower at lower levels of experience
- More experienced locomotive engineers know where to put their brake but a younger engineer might put the brake on earlier
- All trains don't work the same (e.g., different types of brake systems). Experience helps.
- In the winter time a train doesn't stop as fast as in the summer time (so have to start braking sooner)

Impact of Weather/Visibility (e.g. Fog):

- Poor visibility increases the chance that a sign will be missed, or that the locomotive engineer will have less cues to help identify the location of the train
- However, interestingly one of the Focus groups argued that "the worse the conditions are the more you pay attention". Roth note: While this is true, I believe that experience across industries suggests that while attention resources do expand with demand, never-the-less an increase in error is observed as conditions degrade.

Portions of track that are more challenging:

- Steepness of hill – downhill grade– challenges braking skill; trouble if lose brakes
- Uphill stop is easier because gravity is helping
- Less familiar territory:
- Territory that ride on less frequently (less opportunity to develop good mental model of territory)

Specific comments of locomotive engineers on memory demands and value of CBTM in protecting against memory lapses:

- Quote: " One of the hardest things of being an engineer are the things that you have to remember that are specific to this trip." Examples are work crew out there. Can forget that. There is usually a sign but not always.
- Remembering is hard. Have paper (Train Bulletin) but there isn't always a place to put it (in the cab) so where you can see it. So, put it in back pocket.
- In some cases the locomotive engineers (e.g., the locomotive engineer observed during the CBTM test) are able to place the Train Bulletin in front of them and write in the list of temporary speed restrictions in the order in which they will come up so that can follow along, crossing them off as they pass.)
- CBTM helps you to remember:

Slow Orders

Work Orders

- Close Clearances (Note: Not sure that CBTM alerts to these – but a locomotive engineer mentioned it as something that could use help in 'remembering')
- Stop and Flag (E.g., grade crossings where the gate is broken – need to stop and flag) – May forget to do it or may not know where need to stop because the location information provided by the dispatcher is vague (e.g., mile post 57.6; or a description in terms of street names) – locomotive engineers said that they use visual landmarks to orient themselves as to where they are – not mile posts to the tenths and not street names that are beyond their range of vision.
- Train crews tend to know where permanent speed restrictions and blocks limits are; it is the temporary ones that are vulnerable to forgetting.

While this opinion was expressed by the first Focus group participants, the second Focus group participants argued the opposite: More likely to violate permanent speed restrictions than temporary speed restrictions because temporary speed restrictions have more severe consequences. They argued that the worse the conditions are the more you pay attention – your senses are really tuned up.

From a Human Factors point of view there is merit to both arguments. It may be that both types of errors arise and different factors contribute to them – this issue requires further exploration and data collection.

Other Topics of Discussion

Role of Conductor:

- Conductor and locomotive engineer have same Locomotive Displays and radio [but the CBTM display is only on the locomotive engineer's side.]
- Serves as an error catching/recovery mechanism (If the conductor isn't sure that the engineer is aware of something or is under control he will say something (e.g., "Do you got it?) -- (This is consistent with Crew Resource Management Philosophy for helping to catch and recover from errors)
- Provides reminders of temporary speed restrictions, work zones, and end-of authority blocks coming up.
- Handles radio communication – relieving the locomotive engineer of this potential source of workload
- The two operate redundantly – serving to catch and correct each other's errors – for example, they both write down movement authority provided over the radio by the dispatcher.
- "An experienced conductor can help a locomotive engineer (Can say 'You do know you have to stop'). Can keep a young engineer honest."

- It was suggested that it might be beneficial to put the CBTM on the conductor's side to enable him to more effectively play this 'reminder' role (This presumes that the audio alert signal would come on early enough to allow the locomotive engineer to respond and avoid a penalty brake)
- The locomotive engineer can also catch and correct errors made by the conductor – An example was given where the train was in the yard (it was 5:00 AM in the morning so the crew was tired) and the conductor told the locomotive engineer that he had 'Restricted Proceed' signal – when in fact the train was supposed to stop. The locomotive engineer was more experienced and recognized that don't get 'Restricted Proceed' signals in a yard so questioned the conductor and discovered the error.
- The conductor is also the one who manually aligns switches when get to a siding. [This requires that he/she get off the train, align the switch for the siding, wait for the train to go by, then manually re-align the switch for the mainline, and then walk to the front of the train.]

Individual Differences:

- Braking style is very individual. CBTM is trying to standardize braking style – that may be a training challenge – especially for the more experienced locomotive engineers

Memory/Work Aids:

- The locomotive engineer that I observed during the CBTM test, wrote down the temporary speed restriction from the train bulletin onto the front sheet where he writes down movement authorities – so that he would have them visible at all times and would not need to flip back to other pages. He put down the speed restriction speed and milepost where it came into effect. He checked them off as he went past them – therefore aiding him in anticipating the next one to come.

Communication with Dispatchers:

- In the CBTM equipped train that we rode, we noticed that in most cases the dispatcher called the train to give them movement authority before they called the dispatcher to request it. This is different from the 'prototypical' case that is usually described where first the train crew calls requesting authority.
- Also, interview of the Locomotive engineer on the CBTM equipped train indicated that it was not unusual for a dispatcher to call and give authority to move through multiple blocks
- Also, it is not unusual for the dispatcher to tell the Locomotive engineer that he doesn't need to call in to release a block immediately after he/she has passed it, but rather it is OK to wait for the dispatcher to call requesting to know what blocks have already been passed.
- The locomotive engineer indicated that dispatchers do this to level their own workload. When they have a low workload period and they know that a train will be

calling in requesting movement authority, they call the train, – this serves to level workload and increase communication and train movement efficiency. Roth observed similar dispatcher behavior in the study of Amtrak dispatchers.

- Dispatchers also call to find out where the train is (because it is dark territory) for train meets.
- There may be delays in dispatcher answering a locomotive engineer if talking with another train.
- Technical Problems with radio communication:

You answer the dispatcher but he doesn't hear you

Communication with Roadway Workers:

- When want to enter a block that is under roadway worker control (work order) the locomotive engineer has to contact the roadway worker to get permission to enter.
- Roadway workers have small radios with weak signals so can be hard to reach them. Have to sit and wait, or call the dispatcher to see if he/she can reach the roadway worker.
- Impact on roadway worker: CBTM increases safety – benefit for their protection.

Party Line Aspect of Radio:

- Helps to overhear. Helps you to know what other trains are up to.
- Down side is that you have to wait until the radio is free

Other Objects interact with:

- Broken Rails:

locomotive engineer is more likely to detect a broken rail by feeling it than by seeing it. [Most broken rails don't cause derailments.]

By the time you can see it there is typically nothing you can do to avoid it [At 40 miles an hour it can take up to two miles to stop the train.]

- Misaligned switches:

At switches, have targets. The direction of the targets indicates how the switch is aligned.

If have a target is indicating alignment – then more likely to be able to see how the switch is aligned (Engine headlights lights target up)

However if the target is missing or pointed incorrectly (e.g., due to vandalism) then can provide misleading information

Might be able to detect and stop in time some places (e.g., when going up hill) but not others.

- Missing/Obscured Signs:

Advanced Warning boards are supposed to be placed 2 miles early to alert to a slow order – but aren't always there.

Can be missing if the roadway worker has just imposed a speed restriction and didn't have a sign with him to put up – so sign hasn't been put up yet.

- ▪ Missing sign 'is not as unusual as would like' because a roadway worker may not have it with him when he detects a need for a speed restriction – it may take a while to get one and put it in place – this is especially true during the day when roadway worker inspections happen.

Can be obscured by vegetation, by a bridge post

Can be placed at an incorrect spot (e.g., a little past the block limit)

Can be vandalized – Although responders indicated that this is rare on this particular territory.

Can be placed on another subdivision (A locomotive engineer gave an example where there was a work order – but he was coming from another subdivision so didn't have an advance warning board (because it was placed some place else))

If they are present, they are visible at night because they are reflected off of headlights.

- Types of Signs:

Permanent speed restriction signs – diamond

Advance warning board – placed two miles ahead

Work zone – is red

Temporary speed restriction is yellow – square sign

End of temporary speed restriction is green – square sign

Sign at the beginning of a block: white rectangle with the name of the block.

Whistle signs (indicating where the locomotive engineer should blow his whistle)

Time required to brake:

- Depends on length of cars and weight

A coal train can be a mile long

- At 40 miles an hour it can take up to two miles to stop a train
- At 20 miles an hour with a coal train, can still take half a mile at least.

Computer-based Locomotive Control Interface vs. Older analog 'knobs and dials' interface (asked only in the second focus groups of locomotive engineers/conductors who had experience with both):

- Setting up the computers is more time consuming. Mechanical interface was easier to use and more reliable.
- Computer displays are too bright at night

- New employees who have only had experience with computerized interfaces have nothing to compare with.
- One advantage though of the newer Locomotive Cabs is that they have air conditioning and that is a real improvement. Makes you feel better. Gives you a good attitude and allows you to focus.
- Engines are more reliable on the new trains and the air conditioning is nice.

APPENDIX B. SUMMARY OF DISPATCHER INTERVIEWS

Introduction

Observations were made at three different dispatch desks that handled primarily dark territory, including the dispatch desk that handles the territory from Spartanburg, South Carolina to Augusta, Georgia, where the prototype CBTM system was tested. In addition we observed and interviewed a chief dispatcher, whose territory included the territory from Spartanburg to Augusta.

In total we observed and/or interviewed a total of seven railroad dispatchers, one chief dispatcher, two training instructors and two managers of the dispatch center.

This section provides a summary of the main findings of relevance to development of the human factors aspect of the ASCAP model.

Operational Environment

The CSX Dispatch Center contained 42 dispatcher desks. Each dispatcher was responsible for a different portion of territory. Approximately 60% of CSXT territory comprised signal territory. The remaining 40% of territory was dark territory. The majority of dispatcher desks included both signal and dark territory.

CSXT dispatchers communicated with train crews and roadway workers primarily over radio. They also communicated by phone. Dispatchers used a Computer Aided Dispatch (CAD) system to enter and keep track of block authorities.

The dispatchers worked under high workload conditions. They continuously received requests over the radio that they needed to address. They can have several radio towers to respond to that are all flashing at the same time. There was a need to complete each radio transaction quickly in order to meet the demands placed on them. Under these conditions, it was easy to understand how a dispatcher could forget to take an intended action, could make (typographical) data entry errors in the CAD system, or could be ‘only half listening’ during the read-back portion of a radio transaction (and therefore make a communication error) because their attention was shifting to the next task.

Another source of performance challenge was that the poor quality of radio reception. Some radio signals were weak. There were dead spots, static, and individuals ‘stepping over’ each other’s conversations. As a result, it can be difficult to hear and understand what was said over the radio. This had two consequences. First, people were unable to understand what was said on the radio in many cases, and required the listener to ask the speaker to repeat him or herself. This increased the duration of transaction. Second, the fact that the poor signal quality increased the listener’s (dispatcher, locomotive engineer, conductor or roadway worker) reliance on their expectations (context and background knowledge) to help resolve ambiguity. People are likely to hear what they are expecting to hear. This use of expectations is a fundamental aspect of human perceptual systems and happens automatically. However, it can contribute to communication errors. The problem is exacerbated by the fact that in many cases terms may sound alike and therefore be confusable (e.g., train numbers). As a result, individuals (dispatcher, locomotive engineer, conductor or roadway worker) may mishear what is said over the radio.

The CAD system was used to enter, delete, and keep track of block authorities given to trains and roadway workers. It contained a number of features intended to prevent and catch dispatcher errors. For example it has a 'readback' screen. When the dispatcher received a request for block authority over the radio, the dispatcher entered the request for block authority on a data entry screen. After it was entered, a 'readback' screen appeared. The dispatcher was supposed to 'readback' the information on this screen to the person over the radio (e.g., Train crew member or roadway worker) who then repeated the information back. Only when the information was repeated back and confirmed to be correct on the 'readback' screen was the transaction completed and authority granted. In addition the CAD system had checks built in that catch errors. For example, if a dispatcher attempted to give a train authority for a block that was already occupied it would not allow it.

Opportunities for Error/Contributors to Error:

Observations and interviews resulted in a deeper understanding of dispatch center operations and the Computer-Aided Dispatch (CAD) system that dispatchers use to enter block authorities. In particular we were able to identify the most common types of errors dispatchers made and the factors that contributed to those errors.

Dispatchers provided extensive information on the types of errors that they can make and the factors that contribute to those errors. In many cases those errors result in a discrepancy between what is entered in the CAD system and what the receiver of the message over the radio believes is the case. This can happen if:

- The dispatcher makes a data entry error in the CAD system. The dispatcher verbally says the right thing to the person over the radio but enters the wrong thing in the CAD system.
- The dispatcher verbally gives more block authority than he/she enters in the CAD system. A problem can arise if later the same dispatcher or a different dispatcher gives the blocks that were verbally authorized to the first train but not entered in the CAD system to a different train.
- There is a verbal miscommunication over the radio so that what the dispatcher believes he/she has given authority for (or taken away) is different from what the individual on the other end of the radio (locomotive engineer, conductor, roadway worker) believes. In that case the dispatcher will enter in the CAD system and 'read back' from the 'read back' screen what he/she believes to be correct, but it will differ from what the receiver on the other end of the radio hears and writes down on his/her authorization forms.
- The dispatcher fails to inform a locomotive engineer of a temporary speed restriction that came in after the train left.
- The dispatcher allows a train crew to leave a switch in the wrong position but forgets to enter a tag into the CAD system indicating 'Reverse Switch'.

Dispatchers provided examples of the most frequent types of errors and the factors that can contribute to them:

Data Entry Errors:

- Can inadvertently cancel a block authority due to a data entry error (intended to cancel a block authority for an entirely different block of an entirely different train but inadvertently selected the wrong row on a table of block authorities)
- Issue/cancel a different number of blocks in the CAD system than the locomotive engineer believes were issued/cancelled.

Communication Errors:

- Mistakenly believe you are talking with a different train:
    - One example occurs where radio signals carry an unexpected distance and the wrong dispatcher (someone who controls an entirely different territory) receives a call from a train (or roadway worker). As a result, the wrong dispatcher gives authority to a block (e.g., because the names are similar).
- 'Hear' the wrong thing due to noisy radio (static, cut-out) where the brain 'fills in' the missing information based on expectations. For example, a dispatcher can mishear location information. For example, given poor radio quality Mile Post 88.2 and milepost 82.2 are confusable.
- Fail to catch errors made by the locomotive engineer during readback because the dispatcher has moved on to the next task and/or because the dispatcher is also subject to the impact of expectations on perception.[15]
- Locomotive engineer may write down something different from what the dispatcher told him, but repeat back correctly what the dispatcher said. As a result, the CAD systems shows something different from what the locomotive engineer wrote down.

Errors Due to Forgetting:

- The dispatcher can fail to enter blocks into the CAD system for which he has given verbal authority to a locomotive engineer. An example is where a dispatcher wants to give a train authority for a set of blocks that cut across two subdivisions. This is a time consuming process, requiring the dispatcher to first enter the blocks for the first subdivision in the CAD system and do the readback, and then repeat the same process for the blocks in the second subdivision. Sometimes, the dispatcher may give verbal authority for the entire set of blocks while he or she enters the information for the first subdivision into the CAD system, and then go back to enter the authorities for the blocks in the second subdivision after the verbal transaction over the radio is finished. However, if the dispatcher's attention is diverted he or she may forget to enter the block authorities for the second subdivision.
- Dispatcher can forget to call a train crew over the radio to tell them about a temporary speed restriction that was issued after the train departed.

[15] Consider the task of proofreading. It is difficult to catch typographical errors by reading a document, because the mind will skip over repeated words, miss misspelled words, and fill in missing words. Asking the proofreader to slow down is not an effective remedy. Professional proofreaders read a document backwards (bottom to top, right to left) so as not to be influenced by meaning in catching typographical errors.

- Dispatcher can forget to enter a track tag into the CAD system indicating a change in switch position (i.e., the dispatcher allows a crew to leave a switch in the wrong position but forgets to put a tag into the CAD system indicating ‘reverse switch’.).

Coverage

Most errors are caught and recovered immediately before they have any safety consequences. Errors (e.g., data entry errors) can be caught by the individual making the error, they can be caught by the person they are talking over the radio with (e.g., a communication error) or a third party can catch them. A dispatcher mentioned several cases where an error was caught by a third party overhearing a conversation on the radio. For example, one dispatcher described the following incident:

> Dispatcher 1 gave four blocks to locomotive engineer of train 1. He inadvertently failed to enter this into the CAD system. After a shift turnover dispatcher 2 began to give Train 2 the same block in the opposite direction (having no way of knowing that dispatcher 1 gave away the blocks to train 1). Fortunately, the locomotive engineer from train 1 overheard the dispatcher over the radio giving permission to a second train and alerted them of the problem.

This example not only illustrates the vulnerabilities to error but the coverage that can be provided by other people in the system to catch and recover errors.

Input on CBTM

We also obtained feedback on the perceived usefulness of CBTM in improving safety from managers of the dispatch center, training instructors, and dispatchers. Everyone interviewed felt that CBTM was a good idea and likely to improve safety.

Among the points made were that CBTM:

- Would stop a train if the dispatcher has not put in the block authority information in the CAD system (i.e., in cases where due to ‘data entry error’ or verbal misunderstanding between the dispatcher and the locomotive engineer, there is a discrepancy between what was said verbally to the locomotive engineer and what was entered in the CAD system).
- Would stop the train if a switch were inadvertently left in the wrong position.
- Would stop a train if it exceeded a speed restriction (e.g., in cases where the dispatcher failed to communicate a temporary speed restriction.)

Opinion with respect to potential for increased efficiency and economic benefits were mixed.

- Dispatchers saw potential for improved efficiency in track usage if they could obtain more accurate information on the location of trains and high rail cars. Several dispatchers suggested that if the train location information obtained by the GPS system that is part of CBTM were displayed to the dispatchers, it would allow the dispatchers to more effectively manage track usage. It would:
  - Reduce the radio communication associated with finding out where a train was;
  - Allow dispatchers to better estimate when a train is likely to reach a particular block, and improve the quality of track management decisions. For example, if the dispatcher knows

that it will be some period of time before a train will reach a particular block, the dispatcher can give permission for maintenance of way work to proceed in that block.

- The managers interviewed were less sure of the potential economic benefits that could be realized by providing dispatchers with more precise information on the location of trains and high rail cars (e.g., by displaying the train location information obtained from the GPS system that is part of CBTM on the dispatcher's displays). In particular the managers interviewed did not feel that the economic benefits would be great for DTC territory, given the current method of operation.
- The managers interviewed felt that the potential economic benefits would be greater in Traffic Warrant Control (TWC) territory, where blocks have moveable limits. In TWC territory, control points or milepost locations determine limits. They can be changed. If you have movable limits, then you are in a position to take advantage of the added information provided by a GPS system (e.g., you could operate trains closer together.) However, under the current method of operation in DTC territory, where block limits are fixed, they felt that more precise trains location information might not help a great deal. Unless there is a change in the rules of operation, the economic benefits would not be great.

APPENDIX C. DETAILS OF CBTM HUMAN FACTORS QUANTIFICATION ANALYSES

Crew-Caused Exceedances

The analysis of the crew exceedances is presented in Section 3.5 of this report. The recommended distribution for use in ASCAP for the probability of a train crew to exceed its limit of authority is a normal distribution, having a mean value of 3.3 x 10^{-6} per block boundary, and with a standard deviation of 6.8 x 10^{-7}.

Dispatcher-Caused Exceedances

The dispatcher can create conditions where the train crew believes they have a valid authority (based on the verbal communications with the dispatcher) but they are unprotected by the CAD system, which could allow an authority to be issued to another train. Examples of this could occur include the following:

- Errors related to use of the CAD System:
    a. Deleting blocks: The way the dispatcher deletes blocks may result in the wrong number of blocks being deleted and released. As a result, the dispatcher can release more blocks than the train crew gave back. Block releases are not addressed by the CAD readback screen in the way other potential communication-related error forms (e.g., issuing the wrong blocks) are.

    b. CADS workaround 1: CAD flags conflicts with two trains. The dispatcher can override the system by answering questions posed by the system that allow the dispatcher to cope with special conditions. This enables the dispatcher to bypass protection provided by the system.

    c. CADS workaround 2: The dispatcher can only grant or release contiguous sets of blocks. It is possible to get around this limitation in issuing multiple authorities to the same train. It is possible to verbally issue blocks and not use the readback screen.

    d. CADS workaround 3: Helper locomotives are given protection (block authority) verbally. The CAD system is not normally involved.

    e. CADS workaround 4: The dispatcher can issue blocks in more than one subdivision verbally and enter the information later in CADS. The dispatcher can forget to enter information in CADS.

- Train Misrouting
    a. Train misrouted: Dispatcher can misroute trains onto the wrong subdivision (i.e., where the dispatcher has no authority to issue a movement permit). This can happen due to inexperience (new dispatchers). This occurred 3 times in 13 years. Also interference from outside sources (Yardmaster or trainmaster "overrides" dispatcher.)

- Radio Miscommunications

a. Duplicate Engine number: When communicating with locomotives from "foreign" railroads (i.e., when leasing a locomotive), the dispatcher and train crew should use both the railroad name and the engine ID as the train number (initials and number). The communicator can forget to use the foreign railroad name and this could lead to the use of a duplicate train ID. (Engine numbers from the newly leased trains are similar to the old numbers).

b. Similar sounding train ID's number can be confused or misheard.

c. Changing radio channels can result in forgetting to change back to the operating (road) channel. The train crew could miss information from a defect detector.

The following two factors were identified as the most important influences on the likelihood of dispatcher-caused exceedances:

- Workload.
  Factors affecting workload were:
    a. Territory size
    b. Traffic density
    c. Ease with which one can contact train & road crews
    d. Mode of operation -- mixed mode of operation is a problem (sharing signal and DTC).
    e. Through traffic vs. local traffic
    f. Day of week and time of day:
        i. Busy time is the day shift and first 3 hours of the second shift (work peaks)
        ii. Combine desks for 3rd shift
- Radio Communication
  Examples of radio problems were:
    a. Radio communication was identified as a major source of job stress for everyone
    b. Increased communication workload due to very poor radio reception (common to ask for repeats 3-6 times on all calls);
    c. Bandwidth accessibility is reaching its upper limits; switching frequencies to find better reception – too easy not to switch back.
    d. Radio traffic problems are more problematic in Spartanburg-August territory than other territories, but people argued that professionalism compensated for these difficulties.
    e. Changing channels can result in forgetting to change back to operating (road) channel.

Analysis

The workshop attendees agreed that the most immediately relevant databases were the dispatcher disciplinary data and the CSXT operating data from FRA. A review of the disciplinary events indicated that up to 14 events over a period of 2 years and 4 months (1999 to April 2001) could be considered relevant to the scope of this analysis for DTC operations, based on a review by a dispatcher supervisor during the qualitative evaluations. A second review by the workshop attendees indicated that 11 events could be relevant in their opinion. As a result, a flat distribution was created to represent the annual rate of associated disciplinary events of events across the CSXT system, with the range of 4.7 to 6.0 per year and a mean of 5.35/year. The workshop participants next discussed the potential for under-reporting (or at least under-counting—not all events lead to disciplinary actions). The range identified was at least a factor of 2 and as high as a factor of 4, with the most likely factors of 3. This was represented by a normal distribution having a 5-percentile value of 2 and a 95-percentile value of 4, with the mean being 3. Using these values lead to mean estimate of dispatcher-related events of 16.05 events/year.

Using the FRA data for CSXT operations for the period corresponding to the disciplinary database, the total train operations was approximately 219 million train-miles, or 89.7 million train miles/year. Using the analysis in Section 3.4 for the crew-related exceedances, the mean number of train-miles associated with DTC operations was calculated to be 28.9 million train-miles/year. This is the exposure rate for the period corresponding to the 16.05 events/year described above.

The mean rate of dispatcher-caused events is therefore 16.05/28.9 x 10^6 per train mile, which corresponds to 5.5 x 10^{-7} per train mile. Again, based on the analysis in Section 3.4, the mean number of miles/block boundary is 6.34. Therefore, the resulting mean rate is calculated to be 3.5 x 10^{-6} exceedances/block boundary. Based on the Crystal Ball analysis, the best fit for the resulting distribution is a Gamma distribution, with a location parameter of –2.02 x 10^{-6}, a scale parameter of 5.12 x 10^{-8}, and a shape parameter of 7.73 x 10^1.

Overspeeding Events

The CBTM system will alert the train crew if they are overspeeding, as described in Section 3.1. Therefore, we included an analysis of overspeeding events in the scope of this study. Two types of over-speeding events were considered: Permanent and temporary speed restrictions. Examples of both of these are as follows:

- Likely Errors Modes: Permanent Restrictions

    a. Train make-up can play a role (e.g., brake profile)

    b. Weather and temperature can impact braking and visual cues: icing affects ability to brake, fog affects ability to see signs, etc.

    c. Heat orders (Form H) are easy to overlook. Heat orders are issued with train messages – can get an update online about a heat message effective from 1-9pm; easy to forget or overlook if the shift started at 6 a.m.

    d. Lack of experience/familiarity with territory.

    e. Equipment restriction (e.g., for particular cars in the consist)

f. Track topology can contribute to errors. The engineer may increase speed going down a hill so the train will pick up enough speed to go up the succeeding hill. This can occur when a helper locomotive is already part of the consist.

g. Equipment problems in the cab can contribute to overspeeding (e.g., speed indicator). Equipment must be accurate to within 3 mph. (supposed to check once, early as possible, every trip, by checking against mileposts)

h. Fatigue, distraction, & complacency can contribute to speeding errors.

- Likely error modes: Temporary Restrictions

  a. Notification of temporary speed restrictions: they will not show up in the general timetables, only in the train order or bulletin order. There can be hundreds of them, all in paper form – information overload.

  b. Train messages – for subdivision

  c. Train bulletin for whole run

  d. Some temporary speed restrictions are imbedded in an operating rule.

  e. Train messages show information good for that train for that trip only; then they can be tossed.

  f. Train messages & bulletin orders are updated en-route by dispatcher over the radio.

  g. Train bulletins are issued quarterly. They cover a lot of territory (up to 100 mile). They contain a lot of information (100's of pages). These bulletins can contribute to workload because they are difficult to process.

  h. Temporary speed restrictions can last for years. They will not show up in the general timetables; they only show up in the train order or bulletin order.

The most important influence for overspeeding in a permanently restricted section is experience and knowledge of the territory. Most inadvertent overspeeding events were described by the workshop participants as usually occurring in the first 2 years of experience.

The following were described as the most important factors for overspeeding in temporary restrictions:

- Equipment problems – speedometer, brakes, lights, rear-end markers
- Track problems
- Work authorities – track work
- Rolling equipment (high and wide cars)
- Signal indications (transient restrictions)
- Temporary yard speed restrictions (not in scope of this analysis)
- Heat restrictions in train orders or by radio can be very easy to overlook

Analysis

A total of 56 overspeeding events were identified in the train crew disciplinary database over a 4-year period, corresponding to an annual rate of 14 per year, assessed for DTC operations. These overspeeding events were detected and were sufficiently serious to lead to the engineer being decertified by CSXT. The workshop participants described the most likely range for under-reporting of overspeeding events to be in the range from 2 to 4, with 3 being the most likely. This was represented by a normal distribution having a mean at 3 and the 5-percentile and 95-percentile points at 2 and 4 respectively. Therefore, the resulting estimated distribution of overspeeding events, allowing for the underreporting, is a distribution with a mean of 42/year.

In the period corresponding to the disciplinary events, the average number of CSXT train miles (excluding yard operations) was 81.5 million train-miles/year. Using the distribution used in the previous analysis of the relative fraction of DTC train miles to the total, the corresponding annual rate is 26.3 million train-miles/year for DTC operations.

Dividing the distribution associated with the overspeeding events by the distribution associated with the operating experience determines the overall rate per train mile. The result is a distribution with a mean of 1.6×10^{-6} per train mile. However, ASCAP calculates the events on a per-speed-zone basis. Even though there are differences in the causes of overspeeding as described above, discussions with the workshop participants indicated that the rates of exceedance would not be very different for permanent and temporary restrictions. Therefore, for this analysis, both temporary and permanent restrictions could be assessed as equivalent. Based on information provided by ASCAP, there are 36 permanent speed restrictions, and a variable number of temporary restrictions. For the McCormick subdivision, workshop participants estimated the most likely range of speed restrictions between 3 and 5 at any one time. For the Spartanburg subdivision, participants estimated the range between 1 and 3. Discrete (integer) distributions for these two ranges were created, with 5 percent values used for the highest and lowest limits.

The total length of track in the test territory is 120.5 miles. Therefore, the mean number of miles per restriction is the 120 divided by the total number of restrictions (represented by the sum of the permanent restrictions [36] and the distributions of temporary restrictions for the two subdivisions. The mean value of this distribution is 42. The distribution of the number of miles per restriction is therefore 120.5 divided by the distribution of the number of speed restrictions, and has a mean of 2.87 miles. Using this as the mean distance for restrictions, the calculated rate for exceedances per restriction is a distribution with a mean of 4.61×10^{-6} exceedances per speed restriction.

Switches

In this event, a train over-runs a wrongly positioned switch and is at risk for derailing or causing equipment damage to the track. It requires two separate contributing events: the switch left in the wrong (unexpected) position, and the train crew failing to stop before over-running the switch. Both will be analyzed here, since both are represented in ASCAP.

I. Manual Switch in Wrong Position

Scope of Analysis:

- Rule book violation:

    a. For train crew leaving in wrong position is violation of rule book rule 104F;
    b. For dispatchers not protecting the switch in wrong position is violation of rule 539 in the rulebook.

- Trains typically travel around 10 mph into a siding through a switch (maximum speed is 15 miles per hour, but for a train entering a siding, the maximum is 10 miles an hour)
- Taking the diverging route at 3 times the correct speed is a derailment source.
- One concern is that the event may not occur to the first train going over a wrongly positioned switch in the following direction. It is not unusual that the first train would go through and may not even know it has run over a wrongly positioned switch. It is the next train (trailing) that might derail.
- One case to consider is the case where the switch is damaged but the target sign still reads OK.
- On the Spartanburg - Augusta track all switches are manual except for two spring switches and one self restoring switch
- Trains going north take the siding
- Malfunctioning switches (e.g., with a gap – not fully closed) are not included in the analysis

Discussion

What is the chance of leaving the switch in wrong position?

- In general it is not uncommon
- With a 2-man crew that is in the cab (there is no caboose), it means that the conductor can have to walk a long distance (as much as 150 cars lengths) to get back to the cab.
    a. This is time consuming, delaying movement of the train
    b. Also physically strenuous, and for older conductors it may be difficult.
- Sometimes a train crew in a siding will leave a switch behind in the reverse position if they have verbal agreement with the conductor of another train that they will re-align the switch (based on radio communication between the conductors of the two trains.)
    a. Facing train (e.g., Southbound) volunteers to switch it (based on radio communication between the two train crews)
    b. Dispatcher indicates that another train (e.g., trailing north bound train) will be following into the siding
        - It was estimated that approximately a third of the time when a train goes into a siding, another train is following it into the siding
- While it is not the established operating procedure, there appears to be evolving an acceptance of the practice of allowing train crews to re-align switches for each other:

a. Conductors will only leave a switch in the reverse position if they get positive confirmation from the other train crew (via radio communication) that they will re-align the switch for them (the train either following or meeting).

b. Participants indicated that there is a question whether a crew can release a block behind the train with switches not re-aligned.

c. The CSXT rule is that the train should not release a block until the switch is restored. Specifically, rule 104 F allows giving up responsibility to another crew to re-align the switch but cannot give up the block until the switch is switched back to the correct position.

    - For the trailing train (following move), they cannot do it without violating this rule
    - For train coming toward you, the other crew can legally re-align the switch via rule 104 F – but not for trains that are following.

d. The problem is that the first conductor is legally responsible; however, for productivity reasons, they need to rely on their fellow conductors.

e. Also, sometimes things change. Sometimes, the conductor will have to walk all the way back from the cab to re-align the switch and then walk all the way back to the cab (instead of getting off the cab at the location of the switch in the first place)

- Under rule 539, a dispatcher is permitted to let a train crew leave a switch in the wrong position as long as the dispatcher protects it. However, this rule is supposed to be used only in emergencies, and conductors and locomotive engineers indicated that it is very rarely used.

Factors that contribute to leaving the switch in the wrong position:

- Miscommunication (e.g., poor radio reception)
- Conductor-to-conductor communication occurs over the radio. Conductors are not allowed to use a cell phone for communication between conductors
- Distraction – leading to forgetting
- Changing plans (e.g., there is supposed to be a following train, but then plans change)
- Convenience/Productivity (time spent walking):

Bad weather, long walk, poor path for walking

- Mechanical difficulty? Experienced conductors can 'feel' that the switch is not operating properly; they are familiar with failure modes and the way a properly operating manual switch feels. In addition, they must look at the switch points to confirm that they are closed.

Analysis

None of the databases available for the workshop provided useful data for this analysis. Therefore, the inputs from the workshop participants—particularly those with current relevant experience (the engineers and conductors on the route)—were the primary inputs, as follows:

- One individual recalled two cases that occurred in the Augusta – Spartanburg territory where a switch was left in the reverse position in the last 8 to 12 years. One case involved a miscommunication between a roadway crew and a conductor. In his recollection in both cases, the oncoming train discovered it but had enough time to stop. One occurred at night and one occurred during the daytime.
- A second individual indicated that he personally ran over one reverse switch (he stopped and reported it) in his 25-year career.
- It was noted that maintenance records might be under-counted because the maintenance crew can fix a switch (that broke when a train ran over a switch in the wrong position) so it would not need to be replaced.
- It was noted that Certification Data Base cannot be used to estimate how often switches are left in the wrong position, because it is conductors who are responsible for aligning switches, and they are not included in the Certification Data Base.
- It was also noted that since two person crew operation has only been in place for the past 10 years, we can only draw on data and experience from the last ten years of operation.

The workshop facilitators drew a graph for the probability distribution of a switch being left in the wrong position. It represented the probability that a switch would be left in the wrong position once in a period of N, where N was 6 months, 1 year, 5 years, and 10 years. The graph showed the probability at 0 for 6 months then rising and staying flat for 1 year to 2.5 years (interpreted as ‘plausible’) then falling again to zero at 5 years. There was no dissent by the participants, who felt that approximately once per year per switch was reasonable for other segments of track with a similar number of trains.

Based on this distribution and there being 7,098 trains per year traveling through the test territory (see Section 3.4), the distribution of the likelihood of a switch being in the wrong position at the time a train approaches has a mean value of 1.29 x 10^{-4} per train.

II. Train Runs Through Wrongly Positioned Switch

Discussion

- The consequences may depend on whether train is “following” vs. “facing” the switch, though for this analysis, the event of concern is just running through a wrongly positioned switch
- At track speed, the probability will depend on whether the crew can see the target in time to stop. The likelihood will depend on if the train speed is slow enough, the terrain is at the right grade and with the right load, and the crew can see far enough
- The target position is the major cue relied on to tell whether the switch is in the correct position:
    a. If the target looks right (target is green for the switch aligned to the main line and red for being aligned to the siding), the crew will assume that the switch will be normal

    b. The crew can also check the switch position by looking at points on the switches. If the target is in the wrong position (unexpected position), the crew could look at the points to check switch position

- Experience also plays an important role. For example, if the crew knew what trains came before, they may build expectations about what position the switch should be
- Targets can be missing or in the wrong position (different from switch position). Malicious vandalism is a serious problem (they can bend the target intentionally to show the wrong state)
- There are a total of five sidings and therefore 10 switches
    a. Of the 10 switches, the workshop participants estimated that crews would be able to see (at track speed) the targets for 3 southbound and 4 northbound switches (7 of 10) because of local track features and layout
    b. If a train is empty, then the train could generally be stopped in time at these 7 switches – but it will depend on the consist
    c. Most of time, trains going north are empty (75% of time) and would be able to stop
    d. If the train was laden, the participants assessed the chance to stop with a load was 50:50
- Experience is a major factor in ability to stop in time.
- Applying emergency brake can be a derail risk
- Braking decision depends on the crew's judgment of load:
    a. You may not know what is in your load – your braking properly based on information given to you
    b. On a mixed manifest train, the consist information may be incorrect
    c. Ice on the rail could be a cohesion problem
- On a typical trip:
    a. Train can go on 5 sidings on the north bound
    b. On the southbound route, a train would typically need to stop 3 to 4 times (depends on the time of day)

Analysis

Case 1 Track speed

Probability that the crew sees the target in time to stop at a 'stoppable' switch when going at track speed

- The crew are more likely to see the target at night than during the day because they are reflective

- The crew will expect the switch to be in the correct position so they may not notice that it is in wrong position
- The discussion by the workshop participants suggested that the crew would notice the target when not distracted 80% of the time

In order to stimulate discussions and based on the above discussions, the workshop facilitators drew a graph for the probability distribution of probability of failing to stop at a ‘stoppable’ switch. It had the highest probability at .2, and sharply decreased to zero below (at .15) and above (at .3). There was no dissent by the workshop participants. This triangular distribution has a mean value of 0.22.

The probability for failing to stop at the “unstoppable” switches is 1.0, by definition.

Case 2: Slow speed

The train would be traveling at a slow speed (less than 10 miles/hour):

- If it was entering siding
- If it was stopping ‘head to head’ (slowing and prepared to stop)

The percentage of time the train will not stop before running over a wrongly switch in the slow-speed case was considered to be extremely small by participants, say one chance in 10,000.

Work Zones

Scope of Analysis:

Covered in analysis:

- Work performed under Rule 707 – preplanned work authority that appears in the train bulletin
- The error of concern is entering the work-zone area during the restricted times (as specified in the train bulletin) without authority, or being in the area at the start of the specified time

Not covered in analysis:

- Authorizations under Rule 704 are for travel authority or for short-term work and the dispatcher assigns them. A 704 authority would not appear in the train bulletin, so the train crew would not know about it unless the dispatcher tells them. In addition, a train would not be given permission to enter a block where there is a maintenance crew in it with a Rule 704 authority. As a consequence, violating a 704 authority would fall under the category of entering a block without authority (analyzed earlier in Section 3.4).

Discussion

The train crew needs to get permission from the employee-in-charge (EIC) of the work-zone to enter it. This is done by radio. However, some 707’s are an “absolute curfew” with no movement at all permitted.

One situation that can arise is that a train comes to a work-zone before the time when the work-zone authority is activated. In this case, the train is allowed to enter the work zone if the crew believes it can get through the work-zone before the time when the work zone authority is activated.

There are signs for 707 work zones:

- Advanced Warning board (2 miles ahead of the work zone)
- A temporary stop sign (red) placed at start of the work zone.
- Rarely, the signs will be misplaced (Someone mentioned that once the stop sign was placed where the advanced warning board was supposed to be and vice versa)

The train crew:

- First relies on the train bulletin sheet to identify work zones
- Second would see the advanced warning sign (this is considered an important factor)
- Then see the stop sign (the last protection – but they may not be able to brake in time once the crew sees the stop sign).

In some cases if the train crew cannot reach the EIC, then they can contact the dispatcher who then tries to reach the EIC. If the dispatcher cannot reach the EIC, the dispatcher can give the train crew authority to go through but at the "restricted speed"[16] (e.g., in case where the roadway workers were incapacitated or they left the roadway and forgot to give back the track). This happens very rarely. One participant mentioned that there are cases where if you stopped at the stop sign for a work-zone you might be blocking a crossing, thereby violating one rule to comply with another.

Monitoring the Radio:

Normally, train crews are on roadway channel unless they go to talk to the dispatcher on dispatch channel. The train crew speaks to EIC on the roadway channel. The dispatch channel is the one used by the train and the dispatcher to give and receive blocks. The EIC may not be monitoring the dispatch channel—he will be monitoring the roadway channel.

Work gangs listening to radio communications will hear the train crews talking – if they hear a train coming, it can contribute to safety.

Most Likely Error Forms and Factors that Contribute to Errors

1. Multiple "slow" orders in the area and multiple 707 work zones with advanced warning boards in the location (so the train crew may miss one because of the number)
    - Workload and memory load – a lot to keep track of – 3 or 4 707 authorities in a row
    - 707 work zones have been known to overlap. (Overlapping ones do not happen any more, however.)
    - Now it is possible to have them back-to-back, so if there are many at a time crews can miss one

[16] The "restricted speed" is the speed the train can travel at such that it can be stopped within half the distance that the crew can see, and in no case is it to exceed 15 mph.

2. Distraction – e.g., attention shifts to ‘STOP AND FLAG’ – especially if the crew just had a near miss at a road crossing
3. Communication
   - Hardware: Physical characteristics of communication device (e.g., static on the radio; weak radio signal).

     Roadway workers have the poorest radios to communicate with.

     May stop the train at an area with a dead spot

     Could turn the radio down because of high static – so might miss a message

   - Human to Human Communication problems:

The train, in trying to get employee in charge (when having two back-to-back 707 work zones) could inadvertently reach the wrong EIC. It was mentioned that this is something that can happen in some territories but it was considered that it is highly implausible in this CSXT territory because in radio communication, the employee in charge is required to provide their specific foreman name and therefore it is less likely to lead to a miscommunication.

Expectations – “I’m expecting the employee in charge to answer and the person is supposed to say which area.” However, the employees may not be following the communication procedure or may be having problems with quality of the radio.

4. Crew thinks it will clear the affected DTC Block before the work-zone is activated.
   - A train is permitted to enter a work zone prior to it becoming activated as long as the crew believes that they can get through the work zone area before the time when the work zone will become activated.
   - If a train crew thinks it will clear the affected block then it will go in. The crew may discuss it among themselves before deciding to go in.

This assumes that nothing unforeseen occurs. Thus, an unforeseen circumstance that delays the passage of the train would be a contributing factor

If a train crew realizes that they will not be able to make it through the work zone before it times in, the train crew may call the EIC to see if they are really going to start on time or willing to let the train through.

5. Misinterpret location (for less experienced crews)
   - In this territory, it is mainly older experienced train crews. There are mostly veteran engineers. However, 50% of the conductors are younger conductors with less than 2 years experience.
6. Intersecting lines. If a work zone is on one line sometimes, trains that are passing through on an intersecting line may not be aware of the work zone:
   - It may not be in their bulletin and there may not be a sign on their line (because the sign has been placed on the track they are crossing).

- One conductor mentioned that experienced conductors are often in the practice of picking up the bulletins for the intersecting line. However, new conductors may not. It is not required to do so.

7. Misses red (stop) board
   - Same issues and factors as in the case of missing a block boundary sign.

Quantification

Data necessary for this event were not available, nor were the attendees knowledgeable of the likelihood of these events. Workshop attendees suggested using the same fraction as for exceeding DTC block authority

CBTM Applications

The analysis of the CBTM system recognized that it is not the purpose of the CBTM system to change the way crews operate trains from the present DTC system. Rather it is intended to be an "overlay"—that is, it will enhance safety by acting as both a reminder and (ultimately) an enforcer of several of the rules (such as those related to block entry, speed enforcement and work zones entry) that currently rely entirely on manual operations. A summary description of the CBTM system is presented in Section 3.1. The current CBTM system used in the trials is considered experimental and several parameters have not been optimized. One prominent example discussed at the workshop was the amount of time the crew gets a warning indication before the application of a "penalty brake." Current experience suggests that under certain conditions, this can be too short (see the discussion in Appendix A). However, CSXT explained that this and other operational parameters will be examined after the trial period and the time between the presentation of the warning and the application of braking (and other braking algorithms) will be adjusted. Using this assurance, the workshop attendees elected to analyze the use of the CBTM system assuming that these trial problems will be resolved.

Three events were identified in the workshop discussions as requiring quantification:

1. The crew fails to gain control of the locomotive/train following indication of a warning before the penalty brake is applied
2. Train crew over-relies on CBTM (a complacency effect)
3. The train crew enters incorrect consist information into the CBTM system

While a number of different opinions were expressed and evidence offered, there was a general consensus among the workshop attendees that there has not been sufficient experience with the CBTM system to make confident projections of its potential impact on human performance. The CSXT local locomotive engineers and conductors who participated in the human factors quantification workshop indicated that while they have had as much experience with CBTM as anyone, they have only had the opportunity to operate a CBTM equipped train a couple of times each. Further, as mentioned earlier, the version of CBTM that has been field-tested is still in a prototype phase and would be expected to improve substantially prior to actual implementation. As a consequence, experience with the prototype version of CBTM is not likely to be representative of performance of the final production system.

Given that this was the consensus position at the workshop, the general recommendation was to perform sensitivity studies to explore how different assumptions about the impact of CBTM on human performance would affect the results of the CBTM case.

The results for each of the three individual CBTM issues discussed at the workshop are summarized below. In some cases numeric probability estimates were elicited from the workshop attendees. These estimates are presented along with the assumptions that served as a basis for the probability estimates. These probability estimates are recommended as starting points for sensitivity analyses.

Analysis

1. The crew fails to gain control of the locomotive/train following indication of a warning before the penalty brake is applied

In this case, a penalty brake occurs that may have been avoidable.

The workshop participants indicated that the application of a penalty brake, itself has the potential for negative consequences. These include:

- It has the potential to cause a derailment
- It may cause the train to stop at an inappropriate location (e.g., on a grade crossing or on a junction)
- It may cause the train to stop where it cannot restart unaided (e.g., a loaded coal train going up an incline

Because of the potential for negative consequences, it is desirable to avoid unnecessary application of the penalty brake.

There are several reasons why the crew may fail to prevent a penalty brake. For example, in the current trials, the locomotive engineers report that the time to respond seems short (though it was noted that CSXT will re-examine this at the end of the trials, as discussed above). Second, the CBTM system does not recognize that dynamic braking is being applied—only the air brakes—and therefore the engineer may be trying to use one braking system without CBTM "knowing it".

For quantification purposes, the workshop participants decided to assume that the production system would include design changes to avoid some of the limitations of the current prototype CBTM. Specifically, the workshop participants recognized that the CSXT will consider revising the braking algorithm following the test period. At the same time, the workshop participants recognized that there are constraints on the length of time of the warning period because too long a warning period would lead to too many false alarms, and too short a warning period would lead to too many missed opportunities for the crew to prevent a penalty brake.

Given the limited testing experience with CBTM and uncertainty in the final design, the consensus of the participants was that the range of probabilities for failing to respond in time to a CBTM warning to prevent a penalty brake in the final implemented CBTM system was in the range of 0.1 to 0.01. This was represented by a lognormal distribution with its 5th percentile value of 0.01, and its 95th percentile value of 0.1 and is truncated at 1.0. The mean value of this distribution is 0.04.

Given the uncertainty expressed by the workshop attendees, it was recommended to perform sensitivity analyses to explore what the impact would be if the probability of failing to respond in time to a CBTM warning to prevent a penalty brake was higher than 10%.

2. The train crew over-relies on CBTM (a complacency effect).

This case relates to the potentially negative effects of future over-reliance on CBTM (i.e., complacency).

The crew choosing to rely on the effectiveness of CBTM to control the train would be an example of complacency or over-reliance on a control system for which it is not intended. As discussed earlier, such complacency can happen as an unintended consequence of using new technologies. The safety significance of such reliance in this study is that when the CBTM system fails (as inevitably, if rarely, it will), the consequence will be the event for which CBTM would ordinarily provide back protection—be it against over-speeding, exceeding an authority, entering a work zone, or over-running a protected switch.

The probability of interest is: P (crew failure|CBTM believed by crew to be working, when CBTM has actually failed), where crew failure could be exceeding block authority, crossing an improperly positioned switch, or overspeeding, and CBTM failure would be a situation where it appears to be working, but fails to give warning or stop the train.

At the workshop, the consensus was that the experienced crews on the Augusta-Spartanburg run would operate under the philosophy that CBTM should never actuate and that their experience will enable them to avoid nearly all warnings. That is, they will push to the limits, but act early enough to avoid CBTM warnings. Under this operating condition, complacency will not be an issue since there is no reliance on and no regular occurrence of CBTM warnings.

Therefore, for current operation of the Augusta-Spartanburg run, there is no change in operator error probabilities from the base case analyzed (operations without CBTM).

P (crew failure|CBTM believed by crew to be working, when actually failed)

= P (crew failure|CBTM is working properly)
= P (crew failure|no CBTM); i.e., the base cases previously quantified

That is, the probabilities of crew failures (e.g., exceeding block authority, crossing an improperly positioned switch, entering work zones without authority or overspeeding) would be the same with CBTM as in the base case.

Note that, for other runs, other railroads, new crews, and changing operating philosophy, this condition may well change; if so, crew failures to act may be more likely with CBTM than without it. Therefore it is suggested by the analysis team that the ASCAP (or any future similar) analysis perform sensitivity studies by increasing the base case (i.e., without CBTM) human error rates by factors of 2, 5, and 10 and observing the effect on the rates of incidents as modeled by ASCAP. This information about the corresponding increase in incident rates would provide CSXT and FRA with information as how important it is to ensure that complacency does not occur.

3. The crew enters incorrect consist information into the CBTM system

Discussions of this case centered around two situations:

- Intentionally entering incorrect consist information (e.g., to manipulate the CBTM braking profile)
- Unintentional errors (e.g., a data entry error by the crew, or an error in the consist description provided in the paper work given to the train crew)

In principle, entering the wrong consist information could be a way for train crews to effectively prevent the intended operation of the CBTM system without it actually being disabled (a logged event). The consist information is used to calculate the braking distance (and hence the time when the crew should start the braking). In principle, the train crews could fool the CBTM by entering the wrong consist information to delay the system's warning or the automatic braking. This could happen if the engineers felt that the system was making them control the trains too conservatively, for example.

In addition, it would be possible for crews to mistakenly enter the wrong consist information. However, there are several ways for this to be detected or limited. First, the train handling may feel "wrong" to an experienced engineer. Second, when the train passes by the first defect detector, the system would flag a mismatch against the consist entry if the number of cars did not match. Third, the CBTM system has error checks built in to limit the size of the input errors.

As with the other cases considered with the use of CBTM in this section, the workshop participants felt that while there was a potential for incorrect consist entry into CBTM, there was not a sufficient experience base to estimate the probability of incorrect entry. While the participants generally felt that intentionally entering wrong consist information was unlikely (e.g., because management is likely to impose disciplinary action for deliberate manipulation of the consist information), no quantification of the likelihood was provided. As to simple erroneous entry of the wrong data, this was also not analyzed since it was felt that ways existed to detect or prevent a significant unintended error (summarized above).

APPENDIX D. GENERAL INTRODUCTION TO PRA

PRA was developed to examine the risks of rare events, events that do not occur frequently enough for data analysis to provide meaningful information. The first integrated PRA was the Nuclear Regulatory Commission's *Reactor Safety Study* or "Rasmussen Report" (NRC, 1975), which was published even before the term 'PRA' had been coined. The *Reactor Safety Study* introduced the event tree/fault tree methodology that has been so widely used in nuclear power plant and other process industry PRA. PRA implements analysis that breaks down accidents into simpler events for which data exist or that experts have direct experience with. These lower level events are amenable to direct quantification. Over the years, the use of PRA has been expanded beyond the characterization of risk to provide a full understanding of the contributors to that risk. This permits active risk management. It allows us to focus on the most important aspects of risk and to optimize the risk management effort. The following general introduction to the ideas of PRA is taken from previous work of the authors (Bley et al., 1992).

PRA is more than a set of tools for analyzing large systems and calculating a risk parameter. It is a process for understanding the safety status of a facility, identifying contributions of people and specific equipment to safety problems, and evaluating potential improvements. At a deeper level, PRA is really a language for addressing uncertainty in all engineering applications. Our structure for all of PRA is shown schematically in Figure 9 as the set of triplets, $\{S_i, \ell_i, X_i\}$ where S_i describes a particular scenario, ℓ_i is the frequency of that scenario, and X_i is the consequence.

PRA, then, is building the complete list of triplets; i.e. the set of all S_i, ℓ_i, and X_i: $+\{S_i, \ell_i, X_i\}$,. Identifying the full set of triplets requires the analyst to structure the scenarios in a way that is complete and is organized to facilitate the analysis. Structuring the scenarios is both an engineering art requiring experience and a nice sense of analysis, and a process drawing on the techniques of logic modeling and traditional engineering and scientific mechanistic calculations.

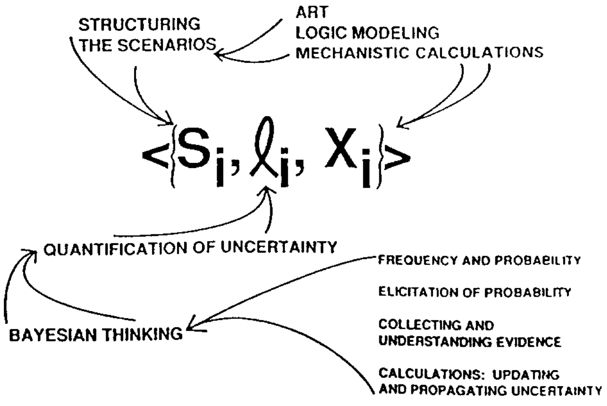

STRUCTURING
THE SCENARIOS
ART
LOGIC MODELING
MECHANISTIC CALCULATIONS
<{Si, ℓi, Xi}>
QUANTIFICATION OF UNCERTAINTY
BAYESIAN THINKING
FREQUENCY AND PROBABILITY
ELICITATION OF PROBABILITY
COLLECTING AND UNDERSTANDING EVIDENCE
CALCULATIONS: UPDATING AND PROPAGATING UNCERTAINTY

Figure 9. The Language of PRA from Bley et al. (1992)

No matter how finely we partition the space of scenarios, it is important to recognize that each scenario really represents a group of similar sub-scenarios. All members of each group must lead to the same consequence. If not, the group should be broken into smaller subgroups until that is the case. The calculation of the frequency of each scenario must be based on considering all possible members of the group; i.e. all possible conditions that might exist under each scenario.

The calculation of the consequences, the X_i, relies on traditional, mechanistic calculations from the engineering disciplines but is distinguished in that consequences from many more cases are calculated than in other approaches. The mechanistic calculations [can] include thermal-hydraulic calculations, electric circuit analysis, … chemical process analysis, and so on. The logic modeling required to structure the scenarios traditionally draws on fault trees and event trees, but other approaches, including digraphs and Markov models, are often used. In some cases, other tools that bridge the gap between logic and mechanistic calculations, such as simulation models, are especially appropriate.

Under the formulation already described, we incorporate the ideas of uncertainty into our calculation of the frequency for each individual scenario group. In addressing the uncertainty of frequency, it is important to adopt a coherent and consistent approach. The Bayesian model provides just such an approach, and under its umbrella, we address the issues of frequency and probability, elicitation of probability, collection and understanding of evidence, and calculations.

Clarity of thought regarding the difference between what we call frequency and probability provides a philosophical framework for understanding a consistent treatment of uncertainty. The two concepts are often confused in the literature of probability, both being called probability. Let us say here that frequency is simply the result of an experiment, be it a real experiment or a 'gedanken' experiment in which we simply count the number of times the event in question occurs out of the total number of possible trials or expired time. Probability, then, represents our state of knowledge about the real world frequency. In the literature, what we are calling probability has gone under various names, including subjective probability, state of knowledge probability, and prevision (deFinetti, 1975; Savage, 1974). Probability, as a measure of what is in our heads rather than a property of the physical world, is a measure of what we know and what we do not know – our complete state of knowledge.

If probability is a personal state of knowledge, how then do we determine probabilities to use in PRAs? Let us consider two cases. In the first case, our state of knowledge comes directly from information that has been collected for other applications; for example, we have collected a wide range of equipment failure data from a variety of power plants around the world.

From these collected data, we have existing curves showing the plant-to-plant variability of, say, the failure rate from motor-operated valves. This plant-to-plant variability curve shows the variation in frequency of failure as we move from plant to plant in a large population. When we now ask, 'What is the probability of failure of motor-operated valves at a new plant? Our probability distribution for the failure rate is numerically identical to the plant-to-plant variability curve or the frequency variability curve.

In other cases, no such plant-to-plant variability curve is available. Therefore, we must elicit the probability from the best experts available to our work. Elicitation of probability is something that is often not done in [PRAs] or not done well. The reasons it is not done well have been documented by (Hogarth, 1975) and others, and include biases built into the human thinking process such as anchoring, overconfidence, and selective interpretation of new data. Careful techniques must be used to avoid these problems.

The last two elements in determining the probability of frequency of each scenario-collecting and understanding the evidence, and calculations using Bayes' theorem for updating probability distributions and propagating uncertainty-are now fairly well established and have been covered in other papers and reports—see for example, Pickard, Lowe and Garrick, Inc. (1983).

The structured language of PRA provides a powerful model for addressing safety and uncertainty involved in all engineered facilities. It provides a framework for organizing a wide variety of standard mathematical and engineering models to address safety issues directly.

APPENDIX E. GUIDELINES FOR HUMAN FACTORS AND HUMAN RELIABILITY ANALYSES

This appendix is intended to provide a set of guidelines for performing a human reliability analysis to insure that the results will be credible, acceptable to the broad set of stakeholders, meet accepted standards for human reliability analysis, and able to be integrated into probabilistic risk assessments.

It is intended to provide guidance for both organizations that are trying to develop an HRA plan as well as regulatory agencies such as the FRA charged with evaluating an HRA analysis that may be submitted as part of a product safety plan.

Four main tasks need to be performed as part of an HRA:

1. *Qualitative Evaluation of Human Factors Issues.* A human factors analysis of the current work environment, the new technology, and their impact on human performance. It requires study of operating rules and procedures and available data, as well as direct observation of the work environment. It allows the analysis team to understand the factors in the current environment that enable errors to be caught and recovered. The goal is to identify the major sources of human risk and reliability with and without the new system. During the qualitative analysis, it is essential that the analysts have direct contact with workers and managers in all aspects of operations. The view from the field is often decidedly different from that held in the central offices; the real-world problems facing operations and maintenance personnel do not always fit the formal procedures and expectations found in design and operations documents or even those found in incident reports.
2. *Survey of databases for HRA sources*. Identify collections of data that may be relevant, problems associated with direct application of that data, and ways in which experts in operations can evaluate and adjust that data to the case at hand.
3. *Quantification*. The process for quantification always begins with an evaluation of the relevance of available data to the actions under analysis. When there are gaps or when the data are not fully applicable to the case at hand, a process must be selected for resolving those issues. That process can involve correctional calculations, telephone conferences with experts in particular areas, small meetings focused on single issues, or a large workshop with all areas of required expertise brought together.

   In many cases the available data bases are insufficient in themselves to support credible human reliability estimation, and the quantification is actually performed during a facilitated workshop that includes experts in PRA and HRA, experts in system design, and people with extensive experience in railroad operations. In this setting, the experts with deep experience in operations examine the models and assumptions to ensure that they represent the system as it is (or will be) operated. Next, experts in analysis and operations jointly examine the available data and agree how it is best used. Finally, for many events there will be no relevant tabulated data, in such cases, the facilitators must elicit the best available evidence from the experience of the experts in operations and design. Together they directly assess the parameters of interest. One advantage of the workshop approach is that all interested parties participate and "buy into" the process and the results. If they are not present or represented, they often resist believing that their points of view were considered. Another advantage is that the evaluators have ready sources of information on issues that affect their evaluations. Still another is that all the evaluators develop a common base of knowledge together. The primary disadvantage is the cost of assembling the group and finding mutually agreeable time. This cost is offset to the extent that review by interested parties not participating in the quantification process and changes that evolve during that review process can be lengthy and difficult.

Whatever approach is used to resolve difficulties in the data, it is important to include a quantification of the range of uncertainty in all estimates, both calculated and directly assessed.

4. *Document process & issues in application*. Finally, to permit review and later understanding of the details of the quantification, all results and processes must be well documented, providing the bases for all estimates. This is especially important and difficult for those cases based on expert judgment: who were the experts; why should we believe them; what were they asked; how was it asked; how did they respond; how were their responses interpreted; and finally did they concur in the analysts' use of their information?

The following are recommendations for how to conduct the HRA tasks that are based on the 'lessons learned' from the present CBTM study.

1. Use an HRA team that includes members experienced in performing human factors studies, human reliability analyses, probabilistic risk assessments and group facilitation techniques
2. Ensure that the modeling of human failure events and unsafe actions is at compatible levels in the PRA and HRA tasks, preferably at the level of available data and experience
3. Ensure that the data sources (databases, expert judgment or a combination) are suitable for the tasks and associated errors that are being analyzed, and that gaps or mismatches are identified and allowed for in the analysis
4. Ensure that qualitative human-factors analyses of the tasks are explored with people experienced in using the systems involved. These should include interviews with workers using the systems (in the case of existing systems) or are the target users of the system (in the case of new technologies that are still under development), and their trainers and supervisors, so that all levels of experience are included
5. The use of expert elicitation methods, when required, should take into account known biases and other limitations of expert judgment. As far as is practical, experts should express their opinions in terms of ranges rather than single point values
6. Inputs should be selected from as broad a range of stakeholders as possible so that the analysis takes into account a wide range of perspectives. However, quantitative inputs should only be accepted during the elicitation process from people with relevant operating experience

7. The results of the analyses should be reviewed by as broad a range of stakeholders as possible to ensure that the broadest possible group will support the results.

REFERENCES

Amalberti, R., & Wioland, L. (1997). Human Error in Aviation. In H. M. Soekkha (Ed.), *Aviation safety: human factors, system engineering, flight operations, economics, strategies, management*. Utrecht: VSP.

Aspinall, W., & Cooke, R. M. (1998). Expert Judgment and the Monserrat Volcano Eruption. Paper presented at the *4th International Conference on Probabilistic Safety Assessment and Management (PSAM-4)*, 13-18 September 1998, New York, NY.

Billings, C. E. (1997). *Aviation Automation: The Search for a Human Centered Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bley, D. C., Kaplan, S., & Johnson, D. H. (1992). The Strengths and Limitations of PSA: Where We Stand. *Reliability Engineering & System Safety*, 38.

Budnitz, R. J., Apostolakis, G., Boore, D. M., Cluff, L. S., Coppersmith, K. J., Cornell, C. A., & Morris, P. A. (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts* (NUREG/CR-6372). Rockville, MD: US Nuclear Regulatory Commission.

Cardosi, K. M. (1993). *An analysis of en-route controller-pilot voice communications* Report. DOT/FAA/RD-93/11. Washington, DC: U S Department of Transportation, Federal Aviation Administration.

Cardosi, K. M., Brett, B., & Han, S. (1996). *An analysis of TRACON (Terminal Radar Approach Control) Controller - Pilot Voice Communications*. Report No. DOT/FAA/AR96/66. Washington, DC: U S Department of Transportation Federal Aviation Administration.

Christoffersen, K. & Woods, D. D. (in press). How to make automated systems team players. To appear in E. Salas (Ed.) *Advances in Human Performance and Cognitive Engineering Research*. Volume 2. JAI Press/Elsevier.

Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York, NY: Oxford University Press.

deFinetti, B. (1975). *Theory of Probability: A Critical Introductory Treatment*. (A. Machi & A. Smith, Trans.). New York, NY: Wiley.

Department of Transportation. (2001). *49 CFR Part 209 et al. Standards for Development and Use of Processor-Based Signal and Train Control Systems; Proposed rule* (Federal Railroad Administration). Federal Register, 66 (155, Friday, August 10, 2001).

Gamst, F. C. (1990). Highballing with Flimsies: Working under Train Orders on the Espee's Coast Line. *The Railway History Monograph: Research Journal of American Railways*, 19 (1 & 2).

Gertman, D. I., & Blackman, H. S. (1994). *Human Reliability & Safety Analysis Data Handbook*. New York: John Wiley & Sons, Inc.

Heath, C., & Luff, P. (2000). *Technology in Action*. Cambridge, UK: Cambridge University Press.

Hogarth, R. M. (1975). Cognitive Processes and the Assessment of Subjective Probability Distributions. *Journal of the American Statistical Association*, 70 (350), 271-294.

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method (CREAM)*. New York: Elsevier Science Inc.

Jordan, B., & Henderson, A. (1995). Interaction Analysis: foundation and practice. *The Journal of the Learning Sciences*, 4, 39-103.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.

Kaufman, L. M., & Giras, T. C. (2000). The Axiomatic Safety-Critical Assessment Process (ASCAP) Simulation Methodology. Paper presented at the *9th IFAC Symposium on Control in Transportation Systems 2000*, Braunschweig, Germany, June.

Kecklund, L. and the project group (2001). *Final report on the TRAIN-project. Risks and proposals for safety enhancing measures in the train driver system*. Borlange, Sweden: Banverket.

Krueger, R. A., & Casey, M. A. (2000). *Focus Groups: A Practical Guide for Applied Research.* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley Publishing Company.

Meyer, M. A., & Booker, J. M. (1991). Eliciting and Analyzing Expert Judgment: A Practical Guide. *Knowledge-Based Systems*, Vol. 5. San Diego, CA: Academic Press.

Monfalcone, M. E., Kaufman, L. M., & Giras, T. C. (2001). Safety Assessment of a Direct Traffic Control (DTC) Train Control System using the Axiomatic Safety-Critical Assessment Process (ASCAP). *Reliability and Maintainability Symposium*, Philadelphia, PA.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. *Human Factors*, 42, 36-55.

Nardi, B. A. (1997). The use of ethnographic methods in design and evaluation. In M. Helander, T. K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed.). Amsterdam, NL: North-Holland.

NRC. (1975). *Reactor Safety Study - An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants* (WASH-1400 (NUREG 75/014)). Rockville, MD: U.S. Nuclear Regulatory Commission.

NRC. (1990). *Severe Accident Risks: An Assessment of Five U.S. Nuclear Power Plants - Final Report*. (NUREG-1150). Rockville, MD: U.S. Nuclear Regulatory Commission.

NRC. (2000). *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis.* (ATHEANA) (NUREG-1624, Rev. 1). Rockville, MD: U.S. Nuclear Regulatory Commission.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors*, 39, 230-253.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics- Part A: Systems and Humans*, 30, 286-297.

Pickard, Lowe and Garrick, Inc. (1983). *Seabrook Station Probabilistic Safety Assessment*. Prepared for Public Service Company of New Hampshire and Yankee.

Reason, J. (1990). *Human Error*. Cambridge University Press, New York.

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Ashgate, Aldershot.

Roth, E. M. and Patterson, E. S. (in press). Using observational study as a tool for discovery: Uncovering cognitive and collaborative demands and adaptive strategies. In Brehmer, B., Lipshitz, R., & Montgomery, H. (Eds.). *How do professionals make decisions?* Hillsdale, NJ: Lawrence Erlbaum.

Roth, E. M., Malin, J. T., and Schreckenghost, D. L. (1997). Paradigms for Intelligent Interface Design. In M. Helander, T., Landauer, and P. Prabhu (Eds.) *Handbook of Human-Computer Interaction* (2nd edition). (pp. 1177-1201). Amsterdam: North-Holland.

Savage, L. J. (1974). *The Foundation of Statistics*. New York, NY: Dover Publications.

Senders, J. W., & Moray, N. P. (1991). *Human Error: Cause, Prediction, and Reduction* (Series in Applied Psychology). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sheridan, T. B., Gamst, F. C., & Harvey, R. A. (1999). *Reliance and distraction effects in PTC automation* (White Paper STD-PTC-DEC-02-99). Cambridge, MA: John A. Volpe National Transportation Systems Center.

Winkler, R. L., & Murphy, A. H. (1968). "Good" Probability Assessors. *Journal of Applied Meteorology*, 7, 751-758.

Woods, D. D., Sarter, N. B. and Billings, C. E. (1997). Automation Surprises. In G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics*, (2nd Ed.) New York, NY: Wiley.

Wreathall, J. (2001). Issues in Selecting a Satisfactory HRA Method. Paper presented at the *OECD/NEA Workshop on Errors of Commission*, Rockville, MD.