Evaluation of the CSA 2010 Operational Model Test

U.S. Department of Transportation
Federal Motor Carrier Safety Administration

August 2011

FOREWORD

In accordance with its primary mission to reduce crashes, injuries, and fatalities involving large trucks and buses, the Federal Motor Carrier Safety Administration (FMCSA) initiated the Comprehensive Safety Analysis 2010 (CSA 2010) Operational Model Test. The program focused on initiating contact with more carriers and drivers, development of a new measurement system to replace Motor Carrier Safety Status Measurement System (SafeStat), application of a wider range of progressive interventions to correct high-risk behavior, and more efficient use of Agency resources. The test ran for 29 months from February 2008 through June 2010.

This report is an evaluation of the CSA 2010 Operational Model Test, focusing on key components of the model.

NOTICE

Technical Report Documentation Page

1. Report No. **FMCSA-RRA-11-019**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle **Evaluation of the CSA 2010 Operational Model Test**		5. Report Date **August 2011**
		6. Performing Organization Code
7. Author(s) **Green, Paul E. and Blower, Daniel**		8. Performing Organization Report No. **UMTRI-2011-08**
9. Performing Organization Name and Address **The University of Michigan** **Transportation Research Institute** **2901 Baxter Road** **Ann Arbor, Michigan 48109-2150**		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address **U.S. Department of Transportation** **Federal Motor Carrier Safety Administration** **Office of Analysis, Research, and Technology** **1200 New Jersey Ave., SE** **Washington, DC 20590**		13. Type of Report **Final Report**
		14. Sponsoring Agency Code **FMCSA**
15. Supplementary Notes **Contracting Officer's Technical Representative: Richard Gruberg**		

16. Abstract

In accordance with its primary mission to reduce crashes, injuries, and fatalities involving large trucks and buses, the Federal Motor Carrier Safety Administration (FMCSA) initiated the Comprehensive Safety Analysis 2010 (CSA 2010) Operational Model Test. The program focused on initiating contact with more carriers and drivers, development of a new measurement system to replace Motor Carrier Safety Status Measurement System (SafeStat), application of a wider range of progressive interventions to correct high-risk behavior, and more efficient use of Agency resources. The test ran for 29 months from February 2008 through June 2010.

This report is an evaluation of the CSA 2010 Operational Model Test. The evaluation focuses on key components of the model. In particular, measurement system thresholds exceeded in type and frequency, interventions received in type and frequency, and the number and percentage of carriers *touched* under the new model are calculated and reported. Intervention cycles and patterns are explored and effectiveness of interventions is determined by comparing test carriers that received CSA 2010 interventions to control carriers that did not. Costs to the agency of performing interventions are investigated and compared to the current process of conducting compliance reviews (CRs). The new Safety Measurement System (SMS) that is used to rank a carrier's safety performance in the seven Behavior Analysis and Safety Improvement Categories (BASICs) is evaluated by assessing associations between percentile scores and crash rates. Comparisons with the current model under SafeStat are reported and quality of the Motor Carrier Management Information System (MCMIS) data files is discussed. Finally, results from a survey completed by field staff participating in the Operational Model Test is presented to determine which aspects of CSA 2010 worked well and which did not.

17. Key Words **CMV, Commercial motor vehicle, crash avoidance, CSA, CSA 2010**	18. Distribution Statement **No restrictions**		
19. Security Classif. (of this report) **Unclassified**	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages **172**	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

Table of APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
In	Inches	25.4	Millimeters	mm
Ft	Feet	0.305	Meters	m
Yd	Yards	0.914	Meters	m
Mi	Miles	1.61	Kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
Ac	Acres	0.405	Hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME	1000 L shall be shown in m^3	
fl oz	fluid ounces	29.57	Milliliters	mL
Gal	Gallons	3.785	Liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		MASS		
Oz	Ounces	28.35	Grams	g
Lb	Pounds	0.454	Kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE	Temperature is in exact degrees	
°F	Fahrenheit	5 × (F-32) ÷ 9 or (F-32) ÷ 1.8	Celsius	°C
		ILLUMINATION		
Fc	foot-candles	10.76	Lux	lx
Fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		Force and Pressure or Stress		
Lbf	Poundforce	4.45	Newtons	N
lbf/in^2	poundforce per square inch	6.89	Kilopascals	kPa

Table of APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
Mm	Millimeters	0.039	inches	in
M	Meters	3.28	feet	ft
M	Meters	1.09	yards	yd
Km	Kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
Ha	Hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	Milliliters	0.034	fluid ounces	fl oz
L	Liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
G	Grams	0.035	ounces	oz
Kg	Kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE	Temperature is in exact degrees	
°C	Celsius	1.8c + 32	Fahrenheit	°F
		ILLUMINATION		
Lx	Lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		Force & Pressure Or Stress		
N	Newtons	0.225	poundforce	lbf
kPa	Kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003, Section 508-accessible version September 2009)

TABLE OF CONTENTS

LIST OF FIGURES AND FORMULAS

LIST OF TABLES

ABBREVIATIONS, ACRONYMS, AND SYMBOLS

Acronym	Definition
BASIC	Behavior Analysis and Safety Improvement Category
CMV	commercial motor vehicle
CR	compliance review
CSA 2010	Comprehensive Safety Analysis 2010
CSP	Cooperative Safety Plan
FMCSA	Federal Motor Carrier Safety Administration
FMCSR	Federal Motor Carrier Safety Regulations
GES	General Estimates System
HMR	Hazardous Materials Regulations
HOS	hours of service
MCMIS	Motor Carrier Management Information System
NOC	notice of claim
NOV	notice of violation
SEA	Safety Evaluation Area
SafeStat	Motor Carrier Safety Status Measurement System
SMS	Safety Measurement System
SSDQ	State Safety Data Quality
USDOT	U.S. Department of Transportation

See the FHWA Terminology and Acronyms supplement for a list of preferred acronyms.

EXECUTIVE SUMMARY

INTRODUCTION

In accordance with its primary mission to reduce crashes, injuries, and fatalities involving large trucks and buses, the Federal Motor Carrier Safety Administration (FMCSA) initiated the Comprehensive Safety Analysis 2010 (CSA 2010) Operational Model Test. The CSA 2010 design includes initiating contact with more carriers and drivers than the prior system, development of a new measurement system to replace the Motor Carrier Safety Status Measurement System (SafeStat), application of a wider range of progressive interventions to correct high-risk behavior, and more efficient use of Agency resources. The major goals of the Operational Model Test are to assess the new design, determine whether it is likely to result in improved carrier safety performance, and to identify any features that need to be adjusted prior to rollout nationally.

This report is an evaluation of the CSA 2010 Operational Model Test. The evaluation focuses on the key components of the Operational Model, including the following elements:

- Evaluation of the Safety Measurement System (SMS) that is used to rank a carrier's safety performance in the seven Behavior Analysis and Safety Improvement Categories (BASICs), including the data used to determine the scores and the associations between the BASICs and crash rates.
- Comparison of the number and percentage of carriers touched by the CSA 2010 process with the existing SafeStat model.
- Evaluation of the effectiveness of CSA 2010 interventions in improving carrier safety in comparison with the SafeStat model.
- Comparison of the costs of performing the different interventions under CSA 2010 with the current process of conducting compliance reviews (CRs).
- Survey of field staff participating in the Operational Model Test to collect their experience and insights on the effectiveness of the CSA 2010 Operational Model.

Originally, the Operational Model Test was conducted in Colorado, Georgia, Missouri, and New Jersey. Carriers in these States were randomly divided into a "test" group that was subject to the provisions of the new CSA Operational Model, and a "control" group that would continue to be monitored by the Agency's current process. For the four original States, the test ran for 29 months from February 2008 through June 2010. Five additional States (Montana, Minnesota, Maryland, Kansas, and Delaware) were phased into the program as *test-only* States at various times as the test proceeded. Since one of the goals of the CSA evaluation was to compare the CSA 2010 process with the existing SafeStat model, the evaluation focused on the test and control groups in the four original States.

SMS METHODOLOGY

The SMS is intended to accomplish four goals:

- Identify unsafe motor carriers for interventions.
- Identify safety problems within broad areas at carriers.
- Monitor the safety performance of carriers on a near-continuous basis.
- Provide input safety measurements to the Safety Fitness Determination process, by which FMCSA identifies carriers that are conditional or unfit to operate.

The SMS consists of seven metrics:

- BASIC 1: Unsafe Driving.
- BASIC 2: Fatigued Driving.
- BASIC 3: Driver Fitness.
- BASIC 4: Controlled Substances and Alcohol.
- BASIC 5: Vehicle Maintenance.
- BASIC 6: Improper Loading/Cargo Securement.
- Crash Indicator.

The data sources for the BASICs are the components of FMCSA's Motor Carrier Management Information System (MCMIS). The MCMIS consists of the Census file, the Crash file, and the Inspection file. Each of these supplies data for the BASIC calculations, either counts of violations or crashes, or exposure measures.

Data from the Census file are used as "exposure" data, to normalize the scores of certain BASICs in order to reflect the exposure of the carrier to crashes or to unsafe conditions. In particular, the counts of power units owned, term-leased, or trip-leased are used as exposure data. Carriers are required to update their information at least every 2 years. Doubtless some of the counts are incorrect, either because the carrier makes an error, does not understand the requirement, or ignores the requirement. However, the analysis and evaluation of the BASICs here have uncovered no reason to suspect a systematic bias.

The MCMIS Crash file is the next primary source of data used as input to the SMS. States are required to report the involvements of trucks and buses in traffic crashes meeting certain severity thresholds: a fatality, an injury requiring transport for immediate medical attention, or at least one vehicle in the crash is towed due to disabling damage. These data are used in the Crash Indicator BASIC, along with counts of power units from the Census file. The Crash Indicator BASIC is most directly related to the actual crash rates of carriers and is most tightly linked to those crash rates.

The primary issues with the MCMIS Crash file related to the SMS are underreporting and reporting latency. In terms of reporting completeness, underreporting is an issue in some States, but comparison with estimates from NHTSA's General Estimates System (GES) shows that the total number of crashes reported to the MCMIS system approximates the number estimated from GES. In addition, FMCSA's State Safety Data Quality (SSDQ) program shows that compliance with reporting requirements is increasing in recent years. In terms of reporting latency, three out of the four test States substantially met the 90-day reporting standard while one did not. Nationally, the FMCSA SSDQ program shows about the same results, with about 70 percent of States rated "good" in terms of timeliness, 24 percent rated "fair" and 8 percent rated "poor."

The final source of inputs to the SMS BASICs is the inspection data, which is collected during roadside checks, traffic enforcement stops, and carrier reviews. Reporting from the test States was very good. All reported more than 95 percent of inspection results within the 21-day requirement.

While the problem of complete crash reporting has not been resolved, there is evidence of improvement. Incomplete crash reporting primarily affects the accuracy of the Crash Indicator BASIC, which is derived from reported crashes. It is a strength of the SMS methodology that multiple indicators are available to identify motor carriers for interventions.

THE RELATIONSHIP OF THE BASICS TO SAFETY AND EFFECTIVENESS IN IDENTIFYING UNSAFE CARRIERS

The SMS was evaluated to determine how well it identifies unsafe carriers. Crash rates were calculated for carriers that exceeded the BASIC thresholds, and were compared to the crash rates of carriers that did not exceed any BASIC thresholds. Crash rates were also calculated for carriers identified under the current SafeStat system and compared to those under the SMS. In addition, scatter plots were made to assess associations between BASIC percentiles and crash rates. To provide a large sample of carriers, crash rates were calculated using 473,847 carriers not participating in the CSA 2010 test, (i.e., carriers from non-test States).

The results showed that the SMS is a significant improvement over the current SafeStat system in identifying unsafe carriers. For all BASICs, crash rates were higher for carriers exceeding SMS thresholds than for carriers not exceeding thresholds. The crash rate was highest for carriers exceeding the Unsafe Driving threshold. Rates were also high for the Fatigued Driving BASIC and the Controlled Substance and Alcohol BASIC. The SMS also identified many more carriers for intervention than did SafeStat. Scatter plots indicate that all of the BASIC measures have positive associations with crash rates, except for two. Excluding the Crash Indicator BASIC, the Unsafe Driving BASIC has the strongest association with crash rates.

Table 1 shows crash rates, calculated over an 18-month span, for 473,847 active carriers according to their BASIC percentile scores determined during February 2008. For each BASIC, carriers were classified as to whether they exceeded that particular BASIC threshold. Note that a carrier may exceed more than one BASIC threshold and may be counted several times among the various carrier groups. For comparison, crash rates are also calculated for carriers that exceeded any BASIC threshold, and those that exceeded no BASIC thresholds. Since those two categories

are mutually exclusive, adding those two categories sums to the total number of carriers in the table. The last column of the table gives the ratio of the crash rate for carriers exceeding each BASIC threshold to the crash rate for carriers that exceeded no BASIC thresholds.

Table 1. SMS Crash Rates Based on SMS Classification February 2008 for Nonparticipating Carriers

BASIC Threshold Exceeded	Carriers	Crash Rate Per 100 Power Units	Ratio to Not Identified
Unsafe Driving	9,245	7.44	3.56
Fatigued Driving	17,959	6.24	2.99
Driver Fitness	3,981	3.04	1.46
Controlled Substance and Alcohol	1,013	6.55	3.14
Vehicle Maintenance	18,700	4.87	2.33
Improper Loading/Cargo Securement	9,409	3.97	1.90
Crash Indicator	5,077	7.32	3.51
Exceeded Any BASIC	44,881	4.94	2.37
Exceeded No BASICs	428,966	2.09	1.00
All Carriers	473,847	3.15	1.51

The crash rates for carriers exceeding BASIC thresholds are significantly higher than for carriers exceeding no BASIC thresholds. The crash rate for carriers exceeding the Unsafe Driving threshold is 7.44, which is greater than the crash rate for carriers exceeding the Crash Indicator threshold, and is 3.56 times greater than the rate for carriers exceeding no BASIC thresholds. Crash rates for carriers exceeding the Fatigued Driving, Controlled Substance and Alcohol, and Vehicle Maintenance BASICs are also high relative to the 2.09 crash rate for carriers exceeding no BASIC thresholds. Also note that the numbers of carriers exceeding the Vehicle Maintenance, Fatigued Driving, Improper Loading/Cargo Securement, and Unsafe Driving thresholds are relatively large. (Note that exceeding the Controlled Substance and Alcohol BASIC is relatively rare.) The total number of carriers exceeding any BASIC threshold is 44,881, or 9.5 percent of all *active* carriers. Using the crash rate as a measure of risk, it appears that the SMS tends to identify unsafe carriers.

For comparison purposes, crash rates were calculated for the same 473,847 carriers based on SafeStat classification. (as shown in Table 2) Carriers are grouped into three SafeStat safety risk categories: SafeStat A/B, SafeStat C, and SafeStat A/B/C. Crash rates are also calculated for SafeStat Not Identified carriers, which are those not classified as A, B, or C. The last column in Table 2 gives the ratio of the crash rate to the Not Identified group. Of the 473,847 carriers that remained nonparticipating during the 18-month followup period, 5,402 are in the A/B category, 3,389 are in the C category, 8,791 are in the A/B/C category, and 465,056 are Not Identified. Table 1 shows that, in the same group, the SMS identifies about five times as many carriers for interventions as the SafeStat system does for the A, B, or C status.

Table 2.SafeStat Crash Rates Calculated Based on SafeStat Classification February 2008

Carrier Group	Carriers	Crash Rate Per 100 Power Units	Ratio to Not Identified
SafeStat A/B	5,402	6.94	2.30
SafeStat C	3,389	4.94	1.64
SafeStat A/B/C	8,791	6.20	2.06
SafeStat Not Identified	465,056	3.01	1.00
All Carriers	473,847	3.15	1.05

The conclusion is that the BASICs are significantly related to underlying carrier safety, though the relationship is less strong for the Driver Fitness and Improper Loading BASICs.

Annually, it is estimated that CSA interventions will touch approximately 6.3 percent of the carrier population (based on all interventions, including warning letters). This compares to about 2.2 percent of carriers that receive full CRs under the current process. Therefore, the number of carriers *touched* by CSA on an annual basis is approximately 6.3/2.2 = 2.9 times greater than the current system based on CRs alone.

Carriers with recent activity are identified based on updates (registration, inspection, review, crash, etc.) to the MCMIS data files, insurance filings, and fees paid to State highway funds (data collected by IRP, Inc.) in the past 3 years. When restricted to carriers with "recent activity", CSA is expected to touch about 9.9 percent of the population. Excluding warning letters, the percentage of carriers expected to be investigated annually under the CSA program is about 6.2 percent. The corresponding percent of "recent activity" carriers with CRs is 3.2 percent. The ratio of carriers identified by the SMS and SafeStat (9.9/3.2 = 3.1) remains fairly constant.

THE EFFECTIVENESS OF CSA INTERVENTIONS

Effectiveness of CSA 2010 interventions was determined by comparing test carriers that received CSA 2010 interventions to control carriers that did not. Three control groups were considered. One control group consisted of control carriers without CRs. In terms of safety, this group was designed to be a low-risk group. Another group consisted of carriers with CRs. This group was designed to be representative of high-risk carriers. The third group was sampled from control carriers to have safety characteristics similar to test carriers prior to evaluation, i.e., these are control group carriers matched to the test group. Carriers were then followed over a 12-month period and evaluated on the percentage of SMS thresholds exceeded.

Interventions are applied as part of a systematic process to change the safety behavior of carriers that exceed BASIC thresholds. In application, it was found that some carriers were subject to only one intervention, while others received a series of interventions. For some carriers, the first intervention was a warning letter, while in others the initial intervention was an onsite investigation or some other intervention. Different combinations of interventions or "cycles" were identified. These cycles are reflective of the safety of the carrier, such that carriers with more BASICs violations had higher crash rates and the problems took longer and required more interventions to resolve.

Table 3 shows the top 10 intervention patterns observed in the test group. There was a total of 79 unique cycles observed. The table also shows the average crash rate for the carriers that fell into each pattern. The most common pattern was carriers that received only a warning letter. These carriers had the lowest average crash rate, indicating that their violations were relatively mild, and it appears that the warning letter—with these carriers—was effective in improving their safety behavior. The second most common pattern was carriers that received only an onsite focused investigation.

Table 3. Primary Intervention Patterns, Cycles Initiated in First Year of Phase II

First Intervention	Second Intervention	Third Intervention	Fourth Intervention	Fifth Intervention	N	Mean Crash Rate
Warning Letter	None	None	None	None	668	2.9
Onsite Focused	None	None	None	None	180	4.9
Warning Letter	Onsite Focused	None	None	None	145	3.6
Onsite Comprehensive	None	None	None	None	130	4.9
Onsite Focused	Cooperative Safety Plan (CSP)	None	None	None	125	4.6
Offsite	CSP	None	None	None	92	4.2
Warning Letter	Offsite	None	None	None	88	3.4
Onsite Comprehensive	Notice of Claim (NOC)	None	None	None	80	6.2
Offsite	None	None	None	None	72	4.1
Onsite Comprehensive	CSP	None	None	None	49	7.5

The next two figures illustrate the results of the interventions for the first two intervention patterns in Table 3 above.

Figure 1 shows the percentage of carriers exceeding at least one SMS threshold for the test group receiving the warning letter only and three control groups. The control group identified as control matched to test is the most appropriate comparison group. These are control group carriers that were carefully selected to have crash rates similar to the test group in the first month of the comparison, as well as to have a similar distribution of the number of BASIC thresholds exceeded. The other two comparison groups are control carriers that received a CR in the followup period and control carriers that did not receive a CR.

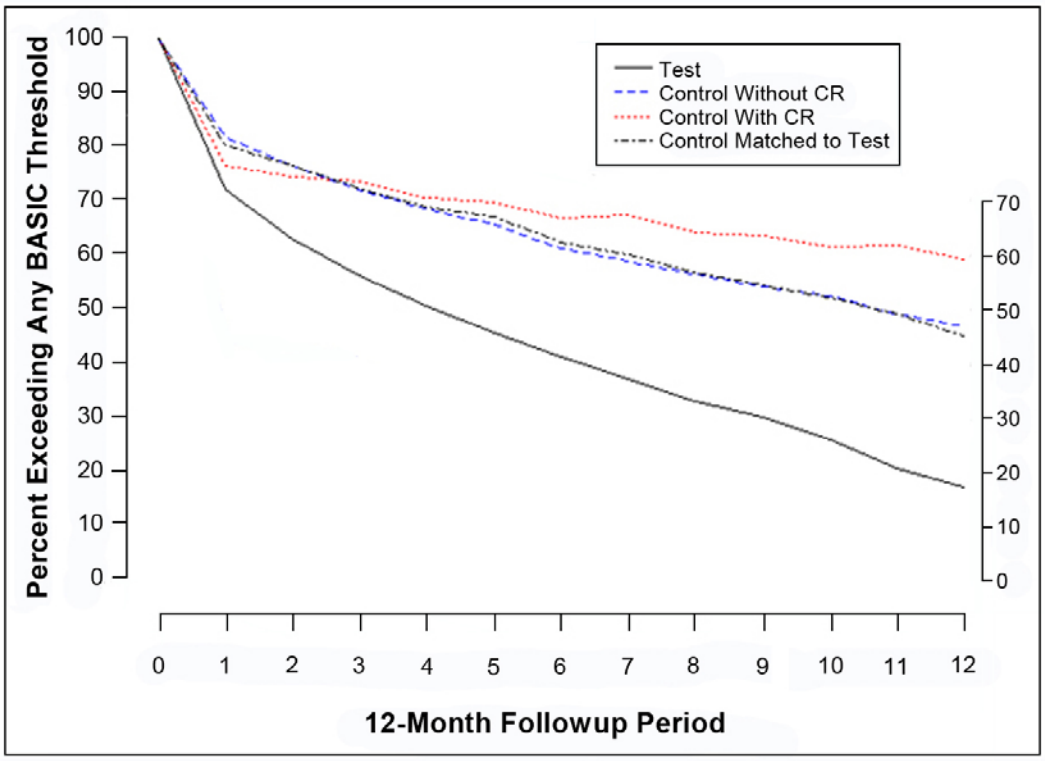

Figure 1. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Warning Letter Only)

The effect of the warning letter intervention is likely one of the most significant findings in this evaluation. After 12 months of followup, only about 17 percent of test carriers still exceeded at least one SMS threshold, compared to about 45 percent of the control group carriers that were matched to test carriers. For carriers receiving CRs during the Operational Model Test, almost 60 percent of carriers continued to exceed at least one SMS threshold after 12 months of followup. Table 3 shows that this intervention pattern is by far the most common, representing about one-third of the total.

Table 2 shows the results for the second most common pattern, carriers that received an onsite focused investigation only. After about 8 or 9 months, the line for the test group crosses lines for both the control group with CRs and the control group matched to the test group. After 12 months, the percentage of the test group exceeding any BASIC threshold is 40.6 percent, which is slightly less than the 43.1 percent for the control group without CRs, but significantly below the control group matched to test group at about 57 percent. Before 9 months, however, the test group has the largest percent of carriers exceeding at least one SMS threshold.

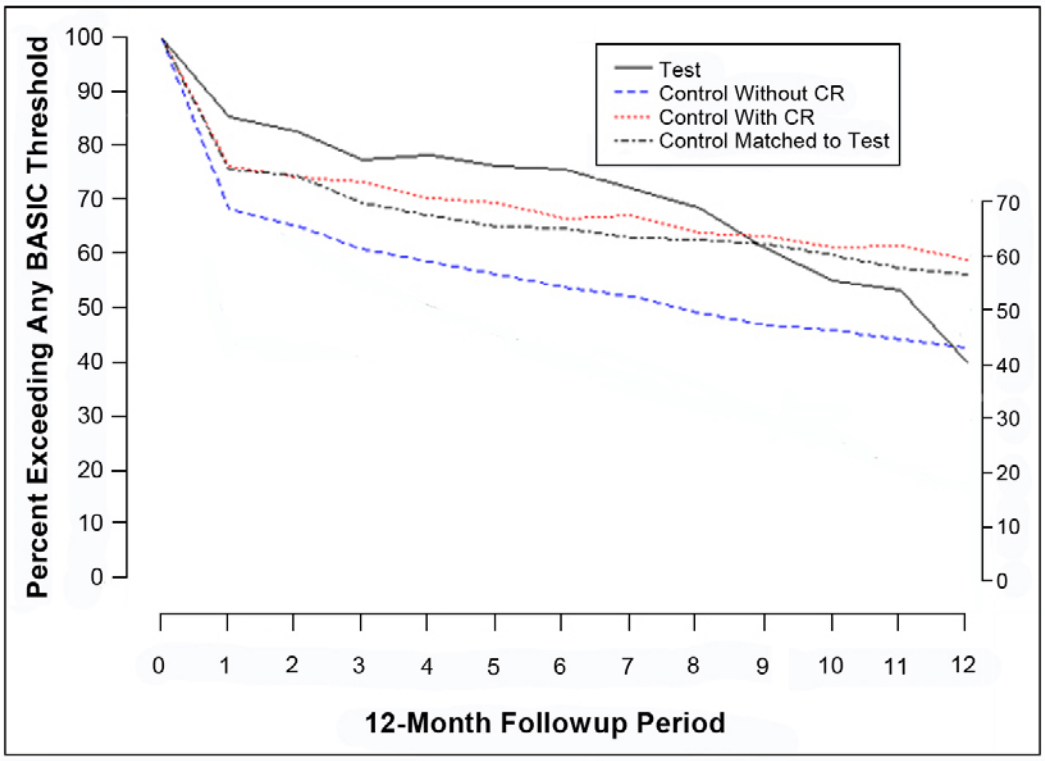

Figure 2. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Onsite Focused Investigation Only)

It appears that the FMCSA was most successful in bringing carriers under SMS thresholds in cases where a warning letter was followed by no other interventions. These carriers tended to have mild violations only, so the warning letter sufficed to resolve the problem. As discussed above, this intervention pattern was also the most prevalent, accounting for about one-third of the patterns and giving the largest sample size of carriers for evaluation.

Effectiveness of any investigation type (offsite, onsite focused, onsite comprehensive) was also examined for carriers with one intervention. A similar pattern emerged. There appeared to be a lag time of about 8 months before carriers with these interventions improved. Once the carriers showed signs of improvement, however, they often had smaller percentages of thresholds exceeded than the control groups. For example, for the onsite focused investigation, after 12 months of followup, 40.6 percent of test carriers still exceeded at least one SMS threshold, compared to about 56.7 percent of control carriers matched to test carriers, and about 60 percent for control group carriers receiving traditional CRs.

For carriers with multiple interventions, it is more difficult to assess the effects that interventions had on safety behavior. For example, intervention patterns that ended in a notice of claim (NOC) generally took about 1 year to be completed. In addition, these carriers tend to be high-risk carriers from the beginning and throughout the investigation process, since interventions were generally escalated to the NOC. Therefore, many of these carriers still exceeded SMS thresholds after the NOC was completed, and the test group did not show lower percentages of thresholds exceeded than a control group, even after matching controls on crash rates and BASIC thresholds exceeded.

COSTS FOR CSA 2010, TEST AND CONTROL GROUPS

Costs to FMCSA of performing CSA 2010 interventions and CRs on the test and non-test groups, respectively, were calculated. All CSA2010 interventions included were those that were classified as closed and completed, to ensure that all tasks performed during a particular intervention have been finished. Table 4 shows the average and median costs per carrier for each intervention type and for CRs performed on non-test group carriers. Warning letters are assumed to incur only a nominal cost, while the other costs include labor hours, travel, and all other expenses. The estimated average total cost per test group carrier that received an intervention is $754, while the median cost is $590. For comparison, the average total cost for non-test group carriers receiving CRs is $1,438 and the median is $1,058. The estimated total annual cost for the test group is $675,000, in comparison with a total estimated annual cost of $785,000 for the non-test group.

Table 4. Adjusted Total Cost in Dollars for Test Group Interventions and Non-Test Group

CSA Intervention Types	Average Cost	Median
Cooperative Safety Plan (CSP)	$95	$72
Notice of Violation (NOV)	$118	$96
Notice of Claim (NOC)	$428	$192
Offsite Investigation	$451	$406
Onsite Focused Investigation	$677	$588
Onsite Comprehensive Investigation	$1,038	$877
Warning Letter	Nominal	Nominal
Estimated Annual Test Group Costs		**$675,000**
Control Group		
CR	$1,438	$1,058
Estimated Annual Control Group Costs		**$785,000**

In addition to the costs to FMCSA of performing interventions on carriers, dollar amounts claimed from test and control group carriers were also calculated. During the 29 month evaluation period of the Operational Model Test, 720 claims were made on test carriers and 640 claims were made on control carriers in the original four states. The average amount per claim was $5,016 for test carriers and $6,296 for control carriers. Because a few carriers had very large claims, the average does not reflect the typical claim, and the median is a better statistic for comparison. For test carriers the median is $2,200 and for control carriers the median is $2,480. The conclusion is that there is not a significant difference in the experience between test and control carriers with respect to claims made.

FIELD SURVEY

A survey of field staff participating in the Operational Model Test was conducted to determine which aspects of CSA 2010 worked well and which did not. The purpose was to collect their insights and perspective on both the process and results of the new system of identifying unsafe carriers and intervening. Responses were received from eight States including all four of the

original test States: New Jersey, Colorado, Georgia, and Missouri. Overall, the evaluation of the CSA was quite positive, both with regard to the process and procedures as well as the outcomes in terms of identifying the right carriers for interventions and having the right tools to improve carrier behavior. There were some critiques of the method of prioritizing carriers for interventions and questions about the quality of the data used to score the BASICs (mostly having to do with the timeliness of the data), but overall the people who were actually operating the new system believe that the new CSA model represents a significant improvement over the prior SafeStat system.

Several respondents noted that more carriers are "touched" by CSA than under the previous regime. They report seeing the same carriers as before but also a large number of new carriers. These "new" carriers are finding out that FMCSA monitors their safety and requires compliance. The field staff respondents generally agreed that the system identifies the right carriers for interventions and provides the right tools to bring carriers into compliance. Several stated, in different contexts that using all available data on safety is the right thing to do.

Several emphasized the usefulness of preserving flexibility in the local divisions. Their approach is clearly predicated on bringing carriers into compliance and they believe that they have the experience, knowledge, and discretion to choose the best tools to do so. Each of the interventions was regarded as useful, depending on the situation. Some were unwilling to vouch for the effectiveness of certain intervention types, specifically the CSP, primarily because enough time had not passed to assess the results. But virtually all clearly indicated that each intervention type could be useful in different situations. It also appears from their comments that most respondents regarded most carriers as willing to comply with the regulations, though there are some carriers that are indifferent to compliance and would not comply without coercion.

Overall, however, the survey of the field staff showed widespread support for the CSA 2010 process. Most would agree that, as one said, "[i]t's better but could still be improved."

1. INTRODUCTION

The primary mission of the Federal Motor Carrier Safety Administration (FMCSA) is to reduce crashes, injuries, and fatalities involving large trucks and buses. In carrying out its mission, the FMCSA has developed programs for monitoring safety performance and compliance with Federal Motor Carrier Safety Regulations (FMCSR) and Hazardous Materials Regulations (HMR) of motor carriers operating on the Nation's highways. The Agency currently collects several types of data on motor carriers, including Federal and State information on crashes, roadside inspections, and enforcement actions. In addition, the Agency uses the data to determine which motor carriers should be selected for onsite compliance review (CR) and to determine the safety fitness of motor carriers. Along with roadside inspections, the CR is the primary tool used for ensuring compliance with FMCSR and HMR.

Until December 2010, FMCSA used the Motor Carrier Safety Status Measurement System (SafeStat) to evaluate the safety status of motor carriers. SafeStat is an automated analysis system designed to incorporate current on-road safety performance information on all carriers with enforcement history information in order to measure relative motor carrier safety fitness. The SafeStat system evaluates carriers in four Safety Evaluation Areas (SEAs): Accident, Driver, Vehicle, and Safety Management. A SafeStat score is calculated that is a weighted combination of the individual SEA values for carriers that meet certain data sufficiency requirements. The scores are ranked and used to prioritize motor carriers for subsequent CRs and roadside inspections.[1]

At current staffing levels, FMCSA and its State partners conduct about 16,000 CRs annually on the approximately 514,000 motor carriers nationwide that FMCSA considers to be active, based on recent activity. Approximately 10,000 of the annual CRs are conducted by FMCSA, while the remaining 6,000 are conducted by State partners. The CR program is resource-intensive and it may take a trained safety investigator several days to complete one CR. Therefore, the program requires considerable Agency and State partner resources and only reaches a small portion of the Nation's motor carriers. After performing a CR, FMCSA issues a safety fitness determination and a corresponding safety rating. One of the limitations of this process is that the safety rating remains in effect until another CR is performed. As a result, a safety rating may not be an accurate indicator of a carrier's current safety fitness. The CR is currently performed as a comprehensive investigation at a carrier's place of business. It has been argued that, for a carrier having deficiencies in one particular area, a full CR may not be necessary. For example, inspections of a carrier over time may reveal a pattern of behavior that is associated largely with vehicle maintenance violations. In such cases, both the agency and the carrier may be better served by a focused investigation designed to address issues specifically related to vehicle maintenance. This would not only relieve the Agency from the costs associated with performing a full CR, but would also allow the carrier to focus on its primary deficiencies.

For these and other reasons, FMCSA has initiated the Comprehensive Safety Analysis 2010 (CSA 2010) program. The goal of the program is to change the current process for monitoring, assessing, and improving the safety performance of motor carriers and drivers through:

- Increased contact with carriers and drivers.

- Development of a new measurement system that replaces SafeStat.
- Application of a wider range of progressive interventions to correct high risk behavior.
- More efficient use of Agency resources.

This report is an evaluation of the CSA 2010 Operational Model Test. The evaluation focuses on the design of the new safety measurement system (SMS), the effectiveness and efficiency of the new intervention processes compared to the current process, the cost to the Agency of applying the new interventions compared to the current process, the safety impact of the new interventions in terms of changes in motor carrier safety performance, and results from a survey of FMCSA field staff to determine their views on the aspects of CSA 2010 that worked well and those that did not. Before exploring each of these specific topics in detail, the next section provides background and design information about the Operational Model Test.

2. THE CSA 2010 OPERATIONAL MODEL TEST

The CSA 2010 Operational Model Test is a field test initiated by FMCSA aimed at implementing components of the CSA2010 program in a small sample of States. It is called an *Operational Model Test* rather than a *Pilot Program* because it did not provide regulatory relief to the carriers participating in the program. For this evaluation, the test began in February 2008, and ended in June 2010, providing a 29-month period for data collection and analysis. Elements of the program are scheduled to be rolled out on a State-by-State or group-of-States basis in 2010 through 2012, subsequent to this Operational Model Test.

In the initial design, the Operational Model Test was conducted in Colorado, Georgia, Missouri, and New Jersey. These States were selected by the CSA 2010 program team based on various logistical, operational, and model design characteristics. For example, each State is from a different FMCSA Service Center; the States are in close proximity to CSA team members; they participate in the Motor Carrier Safety Assistance Program; and these States represent a good cross section of the carrier population. These four States account for approximately a 10th of the Nation's interstate motor carriers and power units.

As the test proceeded, five additional States (e.g., Montana, Minnesota, Maryland, Kansas, and Delaware) were phased into the program at various times. This report focuses mainly on evaluation of the test with respect to the original four States. These States provide the greatest amount of consistent information from the beginning of the test to the end of the test over the 29-month period in which data were collected. Also, carriers in each of the original four States were randomly placed into either a "test" group or a "control" group. Test group carriers participated in the Operational Model Test while control group carriers continued to be monitored by the Agency's current enforcement process. Comparison of the test and control groups based on certain metrics, such as crash rates, helps to assess the effectiveness of the Operational Model Test. The additional five States were added as *test-only* States, meaning that control groups were not formed and all carriers in those States participated in the Operational Model Test as test group carriers. In this report, data collected from the five test-only States are largely used to validate findings from analysis of the test group in the original four States.

The new SMS and the intervention process are two key components of the CSA 2010 Operational Model Test. The SMS is the major tool for measuring the safety of individual motor carriers and commercial motor vehicle (CMV) drivers and is designated to replace SafeStat when the CSA 2010 model becomes fully operational. While SafeStat evaluates carriers in four SEAs, the SMS ranks a carrier's safety performance relative to its peers in seven Behavior Analysis and Safety Improvement Categories (BASICs). A carrier that exceeds a BASIC or Crash Indicator threshold is subject to any of a number of interventions, depending on the number and type of thresholds exceeded. Therefore, the SMS methodology acts as a *trigger* that starts the intervention process. The seven BASICs are briefly described below, followed by a brief description of the CSA 2010 intervention types.

2.1 THE BASICS

Using roadside performance data recorded in the Motor Carrier Management Information System (MCMIS) database, carriers under CSA 2010 are scored in seven BASICs. Unlike SafeStat scores that are used to prioritize carriers for full CRs, BASIC scores are designed to identify specific safety problems which can be addressed by more focused interventions. Carriers are scored in each of the BASICs, ranked relative to their peers, and then assigned a percentile score. Carriers that exceed predetermined thresholds for a particular measure are identified for CSA 2010 intervention. The seven BASICs are described briefly below. They are discussed in more detail in Section 4.

- Unsafe Driving—this score measures the operation of commercial motor vehicles (CMVs) in a dangerous or careless manner. Example violations include speeding, reckless driving, improper lane change, and inattention.
- Fatigued Driving—this score measures the operation of CMVs by drivers who are ill, fatigued, or in non-compliance with the hours-of-service (HOS) regulations. This BASIC includes violations of regulations related to the complete and accurate recording of log books according to HOS requirements and the management of CMV driver fatigue. Instances related to the Fatigued Driving BASIC are distinguished from incidents where unconsciousness or an inability to react is brought about by the use of alcohol, drugs, or other controlled substances.
- Driver Fitness—this score measures the operation of CMVs by drivers who are unfit to operate a CMV due to lack of training, experience, or medical qualifications. Example violations include failure to have a valid and appropriate commercial driver's license and being medically unqualified to operate a CMV.
- Controlled Substance and Alcohol—this score measures the operation of CMVs by drivers who are impaired due to alcohol, illegal drugs, and misuse of prescription or over-the-counter medications. Example violations include use or possession of controlled substances or alcohol.
- Vehicle Maintenance—this score measures the operation of CMVs that are not properly maintained. Example violations include brakes, lights, other mechanical defects, and failure to make required repairs.
- Improper Loading/Cargo Securement—this score measures CMV incidents of shifting loads, spilled or dropped cargo, and unsafe handling of hazardous materials. Example violations include improper load securement, cargo retention, and hazardous material handling.
- Crash Indicator—this score measures the carrier's history of crash involvement, including frequency and severity.

One of the characteristics of the SMS is that calculation of a BASIC measure depends on violations of certain FMCSR and HMR. As part of the roadside inspection program, approximately 3.3 million inspections are conducted annually. Overall, the SMS uses all safety-based violations recorded during roadside inspections. Calculation of the Crash Indicator

depends on the number and severity of crashes recorded during the previous 24 months that are reported by individual States.

Other characteristics that are taken into account when calculating the BASIC measures include violation and crash severity weights, time weights, normalization, peer grouping, and data sufficiency. That is, certain weights are imposed depending on the severity and timing of violations. Violations that are determined to be more severe are assigned larger weights. In addition, the time weight of a violation decreases with time. Normalization refers to the process of accounting for differences in *exposure* among drivers and carriers. Therefore, depending on the BASIC, measures are generally normalized (divided by) power units, driver inspections, or vehicle inspections. Carriers thought to have similar levels of exposure are placed in peer groups and are ranked among their peers. Lastly, carriers and drivers must meet certain data sufficiency requirements before a percentile score is assigned. This ensures that there are enough inspections or crashes to produce meaningful measures of safety.

2.2 THE INTERVENTION TYPES

As part of the CSA Operational Model Test, a test carrier is identified for intervention when it exceeds one or more of the BASIC thresholds as determined by the SMS. The intervention process is designed to be progressive, with subsequent interventions on the same carrier increasing in terms of the scope and intensity of the investigation, until the behavior that precipitated the action is resolved.

- Warning Letter—correspondence is sent to a carrier's place of business that points out which thresholds were exceeded and describes possible consequences of continued safety problems in these areas.
- Offsite Investigation—a carrier is asked to voluntarily submit documents to enable FMCSA to evaluate the carrier's safety management programs and determine the reason for a safety problem.
- Focused Onsite Investigation—an investigation is undertaken at the carrier's place of business when the carrier exhibits a persistent safety problem in one specific area. The investigation is triggered by the carrier exceeding a particular BASIC threshold over a period of time.
- Comprehensive Onsite Investigation—an investigation is performed at the carrier's place of business when the carrier exhibits broad and complex safety problems such as exceeding multiple BASIC thresholds over time. This investigation is similar to the CR conducted under the Agency's current model.

In addition to the above-mentioned types of interventions, the agency may take the following types of action during the course of any investigation:

- Cooperative Safety Plan (CSP)—a safety improvement plan that is implemented by the carrier on a voluntary basis for problems which the motor carrier expresses a willingness

to remedy. The Agency monitors the carrier's safety performance and escalates intervention if performance does not improve.

- Notice of violation (NOV)—a notice to increase the carrier's awareness of enforcement intent on the part of the Agency. Violations discovered are severe enough to warrant formal action but do not warrant fines. To avoid fines or further enforcement, the carrier may need to provide evidence of corrective action or challenge the violations.
- Notice of claim (NOC)—a notice asserting a civil penalty triggered by evidence of severe regulatory violations sufficient to justify assessment of penalties.

2.3 PHASE I AND PHASE II

The CSA 2010 Operational Model Test was implemented in two phases. Phase I began in February 2008, and continued through the end of September 2008. This phase of the test served as a start-up period before the test became fully operational. During this period the test was restricted to the original four States and to interventions arising from three BASICs:

- Unsafe driving.
- Fatigued driving.
- Vehicle maintenance.

Focusing on these BASICs during Phase I allowed the safety investigators in each State to get up to speed with the new CSA2010 protocols more easily. These particular BASICs were chosen because they provide the bulk of the information on carriers exceeding at least one BASIC threshold. Also, during Phase I, carriers having had a CR within the past 18 months, as well as SafeStat A/B carriers, were excluded from the test. In Phase II, the Operational Model Test became fully operational.

3. SCOPE OF THE CSA 2010 OPERATIONAL MODEL TEST

The intervention process begins when a carrier exceeds any one of the seven BASIC thresholds of the SMS. Once that happens, an intervention may be initiated, where a carrier is subject to any of a number of agency actions, depending on which SMS thresholds were exceeded and the number that were exceeded. The carrier is then monitored over time to determine if percentile scores have improved enough to indicate that the intervention process should be terminated, or if percentile scores have not sufficiently improved, the process should be escalated with additional interventions,.

A natural starting point for evaluating the CSA 2010 process is to determine which BASIC thresholds were most likely to be exceeded and the frequency with which they were exceeded. Since interventions result once SMS thresholds are exceeded, the next step is to examine the interventions and intervention patterns carriers were likely to receive. The number of carriers *touched* by the new CSA 2010 model, in comparison with SafeStat, is also assessed. Finally, the effectiveness of the interventions themselves are assessed, by examining motor carrier safety performance prior to and subsequent to each intervention. First, however, crash rates are calculated for the test group and the control group to determine if there were any differences in this measure of safety before the test began.

3.1 CRASH RATES PRIOR TO TEST

Carriers in the original four States were randomly allocated into test and control groups prior to February 2008, when the Operational Model Test began. Ideally, there should be no difference in crash rates between the two groups prior to the test. Table 5 shows crash rates per 100 power units for the test and control groups calculated over the 2-year period between February 2006 and January 2008. The crash rates are 3.72 and 3.32 for the control and test groups, respectively. A rate ratio close to one suggests no difference. It appears the crash rates were substantially similar prior to the test. The control group has slightly more power units, likely due to a few large carriers.

Table 5. Crash Rates for Test and Control Groups Per 100 Power Units Prior to Test (February 2006–January 2008)

Control Group Crashes	Control Group Power Units	Control Group Rate	Test Group Crashes	Test Group Power Units	Test Group Rate	Crash Rate Ratio
7,257	194,935	3.72	5,956	179,272	3.32	1.12

Table 6 shows crash rates further categorized by State. The rate ratios are close to one, even though the one in Georgia is 1.29. Power units counts for test and control groups are also very similar, with the exception of Missouri where the difference is about 16,000 power units. Again, this is likely due to a few carriers with a large number of power units. In general, crash rates prior to the test appear to be reasonably similar between the test and control groups.

Table 6. Crash Rates for Test and Control Groups Per 100 Power Units by State Prior to Test (February 2006–January 2008)

State	Control Group Crashes	Control Group Power Units	Control Group Rate	Test Group Crashes	Test Group Power Units	Test Group Rate	Rate Ratio
Colorado	824	25,330	3.25	779	25,153	3.10	1.05
Georgia	2,334	57,259	4.08	1,835	57,955	3.17	1.29
Missouri	2,466	56,032	4.40	1,720	39,814	4.32	1.02
New Jersey	1,633	56,314	2.90	1,622	56,350	2.88	1.01
Total	**7,257**	**194,935**	**3.72**	**5,956**	**179,272**	**3.32**	**1.12**

3.2 SMS THRESHOLDS EXCEEDED

This section begins the process of identifying the primary patterns of BASIC thresholds exceeded by the test and control carriers. These patterns reflect the underlying safety performance of the carriers, which in turn affects the types and number of interventions applied, and the outcome of those interventions. Subsequent sections identify and evaluate the effectiveness of the interventions in reducing the number of BASICs thresholds exceeded.

When the Operational Model Test began in February 2008, there were 35,008 test carriers and 34,961 control carriers, for a total of 69,969 in the original four States. Table 7 shows the distribution of these carriers by number of SMS thresholds exceeded during February. Overall, 93.9 percent of carriers exceeded no SMS thresholds. Of the approximate remaining 6.1 percent, 4.5 percent exceeded one threshold, 1.3 percent exceeded two BASIC thresholds, and 0.3 percent exceeded more than two thresholds. It appears that the great majority of carriers exceeded no SMS thresholds. (There are data sufficiency requirements for calculating BASIC thresholds. Most carriers do not have a valid BASIC score calculated because of insufficient data. See section 4 for further discussion.) Among those carriers exceeding at least one threshold, 3,114/4,267—or about 73 percent—exceeded only one and the remaining 27 percent exceeded multiple thresholds.

Table 7. Number of Carriers Exceeding BASIC Thresholds by Number Exceeded (Test and Control Carriers, Original Four States, February 2008)

Number of BASIC Thresholds Exceeded	Carriers	Percent
0	65,702	93.9%
1	3,114	4.5%
2	921	1.3%
3	195	0.3%
4	29	<0.1%
5	8	<0.1%
Total	**69,969**	**100.0%**

Table 8 shows the distribution of BASIC thresholds exceeded for the 3,114 carriers that exceeded one threshold. Two BASICs have relatively large and similar percentages. The Vehicle

Maintenance BASIC accounts for 37.4 percent of the total, while the Fatigued Driving BASIC accounts for 35.2 percent. Among carriers exceeding one threshold, about 10.2 percent is attributed to the Improper Loading/Cargo Securement BASIC. The Unsafe Driving and Crash Indicator BASICs have similar percents, 6.1 and 5.8, respectively. The Controlled Substance/Alcohol BASIC threshold is rarely exceeded and accounts for a very small percentage of the total.

Table 8. SMS Threshold Exceeded for Carriers Exceeding One Threshold (Test and Control Carriers, Original Four States, February 2008)

SMS Threshold Exceeded	Carriers	Percent
1—Unsafe Driving	189	6.1%
2—Fatigued Driving	1,095	35.2%
3—Driver Fitness	142	4.6%
4—Controlled Substance/ Alcohol	25	0.8%
5—Vehicle Maintenance	1,164	37.4%
6—Improper Loading/Cargo Securement	319	10.2%
7—Crash Indicator	180	5.8%
Total	**3,114**	**100.0%**

Table 9 shows the different combinations of BASIC thresholds exceeded for the 921 carriers that exceeded two thresholds. The area in the table below the diagonal presents counts of carriers with two BASICs exceeded, while the upper portion presents the corresponding percentages. The BASICs are referenced by numbers 1–7 and cells with large numbers are shaded for ease of identification. Consistent with Table 8, the Fatigued Driving and Vehicle Maintenance thresholds are most likely to be exceeded for carriers with exactly two thresholds exceeded. These two BASICs account for 268, or 29.1 percent of the total. Two other pairs of BASICs account for considerable percentages. The number of carriers that exceeded the Vehicle Maintenance and Improper Loading/Cargo Securement thresholds is 196, or 21.3 percent of the total. The corresponding number for the Unsafe Driving and Fatigued Driving thresholds is 119, or 12.9 percent. Note that the Vehicle Maintenance BASIC is contained in the two pairs with the largest number of carriers, accounting for 29.1 + 21.3 = 50.4 percent, or more than half the total.

Table 9. SMS Thresholds Exceeded for Carriers Exceeding Two Thresholds (Test and Control Carriers, Original Four States, February 2008)

BASIC Measure	1	2	3	4	5	6	7
1—Unsafe Driving	–	12.9	0.8	0.1	3.1	1.8	3.7
2—Fatigued Driving	119	–	4.6	1.3	29.1	3.6	2.5
3—Driver Fitness	7	42	–	0.0	5.4	3.6	1.0
4—Controlled Substance/ Alcohol	1	12	0	–	0.4	0.2	0.0
5—Vehicle Maintenance	29	268	50	4	–	21.3	2.8
6—Improper Loading/Cargo Securement	17	33	33	2	196	–	1.7
7—Crash Indicator	34	23	9	0	26	16	–

Relatively few carriers exceeded more than two thresholds at any particular point in time. Although not all patterns are shown, Table 10 shows the most common patterns for carriers that

exceeded three or four thresholds. (The BASICs are identified by number in this table; their names are shown in Table 9.) These carriers were most likely to exceed the Unsafe Driving, Fatigued Driving, and Vehicle Maintenance BASICs simultaneously. The next most likely pattern involves Unsafe Driving, Fatigued Driving, and the Crash Indicator. In fact, the Unsafe Driving BASIC appears in six of the eight patterns, and the Fatigued Driving BASICs appear in seven of the eight patterns shown in Table 10. It will be shown later that these BASICs have the strongest associations with crash rates.

Table 10. Most Common Patterns of SMS Thresholds Exceeded for Carriers Exceeding Multiple Thresholds (Test and Control Carriers, Original Four States, February 2008)

Number of Thresholds Exceeded	BASIC 1	BASIC 2	BASIC 3	BASIC 4	BASIC 5	BASIC 6	BASIC 7	Carriers	Percent
3	x	x			x			39	29.5%
3	x	x					x	26	19.7%
3		x			x	x		18	13.6%
3			x		x	x		17	12.9%
3	x	x	x					12	9.1%
3	x	x				x		10	7.6%
4	x	x	x		x			5	3.8%
4	x	x			x	x		5	3.8%
Total								**132**	**100.0%**

In summary, it appears that when one or two thresholds are exceeded, the Vehicle Maintenance and Fatigued Driving thresholds are the most prevalent ones. When more than two thresholds are exceeded, in addition to these two, the Unsafe Driving BASIC tends to be included.

3.3 INTERVENTIONS RECEIVED BY CARRIERS

The intervention process is triggered when a carrier exceeds at least one SMS threshold. Table 11 shows the percentage of carriers receiving interventions by number of interventions. A total of 5,587 carriers received 10,281 interventions during the 29 months of the CSA 2010 Operational Model Test in the original four States. The majority of carriers intervened upon, 49.8 percent, received one intervention. An additional 26.9 percent received two interventions and 16.6 percent received three. Among carriers receiving interventions, about 6.9 percent received more than three.

Table 11. Percent of Carriers Receiving Interventions by Number of Interventions, Original Four States, 29 Months

Number of Interventions	Carriers	Total Interventions	Percent Carriers
1	2,780	2,780	49.8%
2	1,501	3,002	26.9%
3	925	2,775	16.6%
4	249	996	4.5%
5	90	450	1.6%
>5	42	278	0.8%
Total	**5,587**	**10,281**	**100.0%**

In this document, when all tasks for a particular intervention have been accomplished and the *intervention* is closed, the intervention will be referred to as *closed/completed.* The closed/completed status does not necessarily imply that the safety *issue* related to the intervention has been resolved. In fact, if the issue has not been resolved, a higher level intervention may be opened. Also, an intervention may be declared closed/completed while the carrier is still being monitored for potential future interventions. However, if a carrier receives one intervention during the early stage of the Operational Model Test, and the intervention is closed/completed without any interventions thereafter, it is likely that the cycle of interventions for that carrier is complete. On the other hand, for example, if a carrier receives a warning letter followed by an onsite investigation and then a CSP, that carrier is still being monitored actively even though the CSP may be declared closed/completed. Nevertheless, closed/completed status indicates that the safety investigator has completed tasks associated with the last intervention recorded.

Table 12 shows the percentage of carriers receiving interventions according to the number of interventions when the last one is closed/completed. Among these carriers, 54.1 percent received one intervention, 22.6 percent received two interventions, 17.1 percent received three, and the remaining 6.2 percent received more than three. Although carriers are still monitored after the final intervention is closed/completed, restriction to carriers with this status provides a basis for the starting point of evaluation forward in time.

Table 12. Percent of Carriers Receiving Interventions by Number of Interventions, Last Intervention Closed/Completed, Original Four States, 29 Months

Number of Interventions	Carriers	Total Interventions	Percent Carriers
1	2,331	2,331	54.1%
2	973	1,946	22.6%
3	737	2,211	17.1%
4	167	668	3.9%
5	64	320	1.5%
>5	33	215	0.8%
Total	**4,305**	**7,691**	**100.0%**

Among the 4,305 carriers whose final intervention is recorded as closed/completed, Table 13 shows the distribution of intervention type by number of interventions. Most carriers received one intervention, and that intervention was a warning letter. Of the 2,331 carriers receiving one intervention, 87.3 percent were warning letters. Investigations (offsite, onsite focused, onsite comprehensive) usually occurred as the first or second terminal intervention. When a carrier with one intervention received one of the investigations, the warning letter was bypassed based on the carrier's safety performance as determined by the SMS. For carriers with two interventions, investigations were likely preceded by a warning letter. The CSP and the NOC are often the final intervention in the series of interventions when they were given to carriers with multiple interventions. Compared to other interventions, the NOV was rarely used. Intervention cycles and patterns are described and presented in more detail in section 4. The purpose here is to identify a group of carriers to evaluate the effectiveness of interventions, since the final intervention is closed/completed.

Table 13. Distribution of Carriers With Interventions by Last Closed/Completed Intervention and Number of Interventions, Original Four States, 29 Months

Last Intervention Closed/Completed	Carriers with 1 Intervention (Percent)	Carriers with 2 Interventions (Percent)	Carriers with 3 Interventions (Percent)	Carriers with >3 Interventions (Percent)
Warning Letter	2,035 (87.3)	89 (9.1)	6 (0.8)	6 (2.3)
CSA	0 (0.0)	268 (27.5)	463 (62.8)	88 (33.3)
NOV	5 (0.2)	21 (2.2)	16 (2.2)	11 (4.2)
Notice of Claim (NOC)	1 (0.0)	98 (10.1)	162 (22.0)	112 (42.4)
Offsite Investigation	84 (3.6)	209 (21.5)	11 (1.5)	3 (1.1)
Onsite Focused Investigation	114 (4.9)	241 (24.8)	56 (7.6)	22 (8.3)
Onsite Comprehensive Investigation	92 (3.9)	47 (4.8)	23 (3.1)	21 (8.0)
Followup Verification	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)
Total	**2,331 (100.0)**	**973 (100.0)**	**737 (100.0)**	**264 (100.0)**

3.4 CARRIERS TOUCHED BY CSA 2010

While there is information recorded in the MCMIS Census file for approximately 750,000 carriers, FMCSA makes a distinction between carriers with "recent activity" and carriers without "recent activity." Carriers with recent activity are identified based on updates (registration, inspection, review, crash, etc.) to the MCMIS data files, insurance filings, and fees paid to State highway funds (data collected by IRP, Inc.) in the past 3 years.

Annually, it is estimated that CSA interventions will touch approximately 6.3 percent of the carrier population (based on all interventions, including warning letters). This compares to about 2.2 percent of carriers that receive full CRs under the current process. Therefore, the number of carriers *touched* by CSA on an annual basis is approximately 6.3/2.2 = 2.9 times greater than the current system based on CRs alone.

When restricted to carriers with recent activity, the percentage expected to be touched annually by interventions is about 9.9 percent. Therefore, one would expect the CSA program to touch approximately 10 percent of such carriers annually when fully implemented (see Table 14).

Table 14. Annual Percentage of Test Group Carriers with Recent Activity Touched by Interventions (Original Four States, 29 Months)

Average Number of Test Carriers With Recent Activity	Total Interventions	Total Carriers With Recent Activity With Interventions	Annual Number of Carriers With Recent Activity With Interventions	Annual Percent of Carriers Touched
22,586	10,095	5,419	2,242	9.9%

Table 15 shows the results presented in Table 14 broken out by intervention type. The three mutually exclusive groups include warning letter only, warning letter followed by additional intervention(s), and all else. The *all else* group consists almost entirely of carriers that initially bypassed the warning letter and received some type of investigation (off-site, onsite focused, onsite comprehensive) as the first intervention. Carriers that received a warning letter as the only intervention represent 37.6 percent of all carriers intervened upon during the Operational Model Test. An almost equal percentage of carriers received the warning letter followed by additional interventions. Carriers that bypassed the warning letter and were escalated to an investigation represent 26.7 percent of carriers with interventions. Excluding warning letters, the percentage of carriers expected to be investigated annually under the CSA program is about 6.2 percent.

Table 15. Annual Percentage of Test Group Carriers With Recent Activity Touched by Interventions by Intervention Type (Original Four States, 29 Months)

Intervention Type	Total Interventions	Total Carriers With Recent Activity With Interventions	Percent of Total Carriers	Annual Number of Carriers With Recent Activity With Interventions	Annual Percent of Carriers Touched
Warning letter only	2,035	2,035	37.6%	842	3.7%
Warning letter followed by additional intervention(s)	5,235	1,936	35.7%	801	3.5%
All else	2,825	1,448	26.7%	599	2.7%
Total	**10,095**	**5,419**	**100.0%**	**2,242**	**9.9%**

From the MCMIS Census file, there are approximately 514,000 carriers with recent activity in the 50 States and the District of Columbia. Of these carriers, 16,262 (3.2 percent) received full CRs in 2009.

The ratio of the percentage of recent activity carriers touched by CSA 2010 interventions to the percentage of recent activity carriers with CRs on an annual basis is 9.9/3.2 = 3.1. It appears that CSA 2010 interventions will touch approximately three times the number of carriers with full CRs annually.

4. THE SAFETY MEASUREMENT SYSTEM

The SMS is a methodology intended to measure the safety of motor carriers. The SMS is intended to accomplish four goals:

- Identify unsafe carriers for interventions.
- Identify safety problems within broad areas at a motor carrier.
- Monitor the safety performance of carriers on a near-continuous basis.
- Provide input safety measurements to the Safety Fitness Determination process, by which FMCSA identifies carriers that are conditional or unfit to operate.

The SMS consists of seven metrics:

- BASIC 1: Unsafe Driving.
- BASIC 2: Fatigued Driving.
- BASIC 3: Driver Fitness.
- BASIC 4: Controlled Substances and Alcohol.
- BASIC 5: Vehicle Maintenance.
- BASIC 6: Improper Loading/Cargo Securement.
- BASIC 7: Crash Indicator.

BASIC 1, Unsafe Driving: This metric is based upon reported moving violations; it is intended to capture driving a CMV in a dangerous or careless manner. The violations captured are all traffic violations, produced by a traffic stop by an enforcement officer. They include typical traffic violations that any driver is subject to, such as speeding, reckless driving, unsafe lane change, or unsafe turn. Since this BASIC is based on observed traffic violations, the violations counted in BASIC 1 are not uncovered during routine inspections, as is the case with several other BASICs.

BASIC 2, Fatigued Driving: This metric incorporates violations of FMCSR governing HOS, completion of log books, and carrier fatigue management practices. Some of the violations are identified during routine roadside or carrier inspections. For example, if a motor carrier enforcement officer makes a traffic stop, he or she may identify violations for this BASIC while reviewing the driver's log book. Incomplete duty status records and a variety of other paperwork violations contribute to this BASIC.

BASIC 3, Driver Fitness: This BASIC captures a large number of violations that reflect a driver who is unfit to drive because of a lack of qualifications, training, certification, or experience. Violations include improper licensing (such as failure to possess a Commercial Drivers License or the appropriate endorsement to operate certain vehicles), the lack of required medical certification, or a carrier allowing an unqualified driver to operate. Data for this BASIC are accumulated during routine roadside inspections, traffic enforcement stops, and carrier inspections, such as occur as part of a CR.

BASIC 4, Controlled Substance and Alcohol: This score captures drivers cited for operating a CMV impaired by alcohol or illegal drugs, or who are found to have misused prescription or over-the-counter medications. Typical examples of violations captured in this BASIC include use or possession of controlled substances or alcohol. The BASIC also includes FMCSR governing the motor carrier's drug and alcohol testing program.

BASIC 5, Vehicle Maintenance: This BASIC is based on roadside inspections, carrier reviews, and post-crash inspections. It captures violations of FMCSR that set standards for the mechanical condition of the different systems on trucks and buses. Examples include brake adjustment, lighting system and conspicuity markings, tire inflation and tread depth, and a variety of other mechanical systems. Paperwork violations, such as failure to maintain maintenance records and failure to document required inspections correctly are also included.

BASIC 6, Improper Loading or Cargo Securement: This basic captures violations of FMCSR that cover loading and securing cargo. Many of the requirements are related to the transport of hazardous materials, including regulations governing appropriate tank types, properly marking the tank or package, appropriate package or tank types, cargo release in a crash or while loading/unloading, appropriate documentation, and appropriate handling procedures of the hazardous material. Other violations included in the BASIC cover proper physical securement of cargo such as through the use of tiedowns or other means of securing the cargo. This BASIC also includes FMCSR governing routing restrictions for hazardous materials, as well as associated planning and documentation. Most of the data used in calculating the BASIC are collected during inspections, whether at the roadside or at the carrier. Violations uncovered during post-crash inspections are also included.

BASIC 7: Crash Indicator: This metric is the most direct indicator of safety, at least in terms of safety being measured by crash involvements. This indicator is based on crash counts from the MCMIS Crash file, normalized by a measure of exposure to produce a crash rate. This measure is the most direct measure of carrier safety, insofar as crash rates reflect the safety of the operations of a carrier.

Nominally the BASICs cover the gamut of factors that can contribute to crashes: driver, vehicle maintenance, carrier operations. Carrier factors as such are not the focus of any specific BASIC, but carrier behavior and conduct are reflected in most of the BASICs. Moreover, virtually all of the factors that figure into the calculation of the BASICs are covered in the FMCSR.

Crash rates are in a sense the best measure of a carrier's safety because reducing crashes is the fundamental goal of improving safety. However, using crash rates as the sole measure of safety presents practical problems. Crashes are relatively rare and stochastic (probabilistic) events, such that whether they occur or not is not entirely determined by the behavior of the carrier and driver. Many factors can contribute to a crash, in addition to poor performance by the driver or vehicle. These include the other drivers and vehicles on the road and the environment itself. Thus, poor and unsafe carriers may have a low crash rate for a period of time, simply because not enough crashes had occurred by chance. This is particularly true of small carriers, which have low exposure to the chance of being involved in a crash, simply because they are small. Over time, poor carriers will have proportionately high crash rates, but it is highly desirable to identify poor

carriers as soon as possible, without waiting for a statistically reliable number of crashes to occur, so that they may be made safer or removed from operation.

Since crash rates may in the short run be unreliable indicators of unsafe operations, alternatives that use events that occur more frequently and that are associated with crashes may be superior. The BASICs as defined in CSA 2010 are designed to serve this end. They function in effect as surrogates for the crash rates.

Whether the BASICs are useful surrogates for crash rates, and by extension, carrier safety, is a testable hypothesis. Presented below is a statistical evaluation of whether the BASICs are related to crash rates, and the strength of the relationship. But first is a discussion and qualitative evaluation of the elements of the SMS, including the sources of the data, the use of weights for violations by severity and how recently the violation occurred. The measures are normalized by exposure, comparison within peer groups, and data sufficiency requirements.

The general process or methodology is based on certain concepts, assumptions, and conditions that affect the ability of the SMS to identify unsafe carriers. The next section provides a discussion of those concepts and assumptions. A rigorous, quantitative evaluation of some of the assumptions is not possible because the data to support such a quantitative evaluation do not exist. For example, one of the implicit assumptions is that applicable violations that occurred within the past 6 months should be given three times the weight of an applicable violation that occurred between 12 and 24 months ago. A quantitative analysis of whether three is the correct weight or 6 months is the correct period is not attempted and is outside the scope of what is feasible in the evaluation. Instead, the authors discuss the elements and assumptions of the methodology in terms of strengths and weaknesses. The true test of the SMS is its ability to accurately identify unsafe carriers. That topic is taken up after the discussion of the elements.

4.1 SOURCES OF DATA

A primary concern for the SMS is the sources of data used to determine the BASIC scores. The question with respect to data sources is whether they are appropriate, the limitations of the data sources, and whether there are alternative sources of data.

The data sources for the BASICs are the components of the MCMIS. The MCMIS consists of the Census file, the Crash file, and the Inspection file. Each of these supplies data for the BASIC calculations, either counts of violations or crashes, or exposure measures.

The Census file contains one record for each motor carrier and shipper registered with the U.S. Department of Transportation (USDOT). Entities that operate trucks and buses in interstate commerce, and all hazardous materials carriers and shippers, are required to register with the USDOT. Most of the information in the Census file is supplied by the carrier or shipper. The information includes contact information, as well as information about the type of operations and the type of commodity carried. There is also information about the number of power units (vehicles), number of drivers, and the amount of travel (vehicle miles of travel [VMT]). The carrier is required to update the information every 2 years (49 CFR 390.19). The data can also be updated if there is a safety, compliance, or educational contact review.

Data from the Census file is used as "exposure" data, to normalize the scores of certain BASICs in order to reflect the exposure of the carrier to crashes or to unsafe conditions. In particular, the counts of power units owned, term-leased, or trip-leased are used as exposure data. Information about the number of power units operated by a carrier is supplied by the carrier itself. It is of unknown accuracy until there has been a CR or some other safety contact. Power units are used as exposure in the Crash Indicator and Unsafe Driving BASICs, so it is important that the number be accurate. In addition, the number of power units from the Census file is used to assign carriers to peer groups. Carriers are then given a percentile ranking for a BASIC within the peer group.

The power unit information is likely quite accurate after a CR, though there is no audit of the data prior to a CR. Carriers are required to update their information at least every 2 years. One source of inaccuracy can be a change in the number of vehicles operated related to business conditions. Doubtless some of the counts are incorrect because the carrier makes an error, does not understand the requirement, or ignores the requirement. However, the analysis and evaluation of the BASICs here has uncovered no reason to suspect a systematic bias.

It must be recognized that there is no real alternative to using the Census file for exposure or normalization. The Census file is the only data source that covers all motor carriers in interstate commerce and all hazardous materials carriers. Thus, there is no alternative for counts of vehicles at the carrier level.

The MCMIS Crash file is the next primary source of data used as input to the SMS. States are required to report the involvements of trucks and buses in traffic crashes meeting certain severity thresholds, namely, a fatality, an injury requiring transport for immediate medical attention, or at least one vehicle in the crash towed due to disabling damage. The data used from the Crash file in the SMS is limited to the counts of crash involvements for carriers and the severity of those crashes. These data are used in the Crash Indicator BASIC, along with counts of power units from the Census file. The Crash Indicator BASIC is most directly related to the actual crash rates of carriers and is most tightly linked to those crash rates.

The primary issues with the MCMIS Crash file related to the SMS are underreporting and reporting latency. Underreporting occurs when crashes that meet the MCMIS Crash file criteria are not reported by the States to the Crash file. Reporting latency refers to the time between when a crash occurs and when it is reported to the MCMIS Crash file. States are required to report all crash involvements that meet the MCMIS reporting criteria within 90 days of the crash. For both of these issues, the effect of the issue is incomplete data used in calculating the Crash Indicator BASIC. (Of course, unsafe carriers could still be detected by other BASICs, which rely on other sources of data, including the MCMIS Inspection file.) Incomplete data may result in unsafe carriers going undetected by the Crash Indicator BASIC. The effect of underreporting and reporting latency on identifying unsafe carriers in the SMS is considered in section 5.5 below.

Underreporting has been a persistent issue for the MCMIS Crash file, which FMCSA has invested extensive resources in addressing. The researchers have conducted a series of evaluations of State data reporting to the MCMIS Crash file. The reports have consistently shown significant underreporting of crash involvements to the MCMIS file, though with an overall trend toward improvement. In most States evaluated thus far, more serious crashes tend

to be reported at a higher rate than less serious crashes, and large trucks, such as tractor-semitrailers and tractor-double trailer combinations, tend to be reported at a higher rate than smaller trucks and buses.

Reporting from three of the four original test States has been evaluated as part of this series. (Colorado has not yet been evaluated.) It should be noted that the most recent report available for Georgia covers crashes in 2006; the evaluation of MCMIS file reporting from Missouri was for the 2005 data year, and for New Jersey, 2003. Table 16 shows some of the results for these State evaluations. Reporting rates are calculated by examining the State's crash file to identify involvements that meet the MCMIS reporting criteria and matching those involvements to the records reported to the Crash file, and then calculating the proportion of records that were reported to the number of records that should have been reported. The overall reporting rate—that is, the rate for all reportable crashes—was 68.1 percent for Georgia 2006, while it was somewhat higher for the Missouri 2005 cases and New Jersey 2003 cases at 83.3 percent and 82.5 percent, respectively. For both Georgia and Missouri, fatal crashes were reported at a substantially higher percentage than the overall reporting rate. New Jersey was somewhat anomalous, with only about two-thirds of fatal crash involvements reported, compared with the 82.5 percent overall rate.[2,3,4]

Table 16. Reporting Rate for Fatal and All Reportable Crashes

State	Evaluation Year	Fatal Crashes Rate	Overall Reporting Rate
Georgia	2006	78.8%	68.1%
Missouri	2005	94.6%	83.3%
New Jersey	2003	67.4%	82.5%

More recent information on the crash reporting from the four test States is available from FMCSA's State Safety Data Quality (SSDQ) evaluation program. States are rated monthly on, among other items, the completeness of reporting to the MCMIS Crash file. In terms of crash file reporting, States are rated separately on the completeness of reporting fatal and nonfatal crash involvements. As of the end of 2010, all four test States were rated "good," the highest rating, on the completeness of their nonfatal and fatal crash involvements. This indicates that both Georgia and New Jersey improved their reporting from the years in which they were evaluated by the researchers. A "good" rating for fatal involvements is assigned to States reporting 90 percent or more of their fatal involvements. The number of fatal involvements reported is compared to the count from the National Highway Traffic Safety Administration's Fatality Analysis Reporting System file, which is the standard census file for fatal traffic crashes. The standard for the "good" rating for nonfatal crash involvements is within 75 percent of the number predicted by a statistical model, using the number of fatal involvements. The prediction method was developed by the researchers.[5] The method assumes that fatal crash involvements are the most reliably identified crashes and that the ratio of fatal to nonfatal reportable involvements is reasonably fixed.

Considering all States, as of the end of 2010, 44 States were rated "good" on fatal crash involvement reporting, four as "fair," and one as "poor." Two States were not rated due to insufficient data. In terms of the nonfatal crash completeness measure, 34 States were rated

"good," 12 were rated "fair," and three were rated "poor," with two States again not rated due to insufficient data.

A third method to gauge reporting to the MCMIS Crash file uses the National Automotive Sampling System General Estimates System (GES) file. These data can be used to estimate the total number of truck and bus crash involvements that meet the MCMIS reporting criteria nationally (i.e., a fatality, injury transported for treatment, or a vehicle is towed due to disabling damage). It is not possible to break out GES data by the individual States, but it is possible to estimate the total number of reportable involvements using GES and to compare that to the total number reported in MCMIS. Table 17 shows the number of involvements reported to MCMIS, the number of reportable cases estimated from GES along with the standard error of that estimate, and the ratio of MCMIS reported cases to GES estimated reportable cases. The ratio shows that the number of cases in MCMIS converged on the GES estimate in 2005, but in subsequent years actually exceeded the GES estimate. Twice the standard error gives the 95 percent confidence interval, so the number of cases reported in each of the years is within the 95 percent confidence interval for the estimate. However, the confidence interval is relatively wide, so this is a weak test of reporting completeness. Moreover, this approach is only able to assess the overall completeness of reporting to the MCMIS Crash file, and does not address any bias that may accrue due to underreporting from a particular State.

Table 17. MCMIS Crash File Cases and Estimated Reportable Cases Using GES

Crash Year	MCMIS Reported Cases	Estimated MCMIS Reportable Cases From GES	Standard Error of GES Estimate	MCMIS/GES
2003	136,418	160,000	±11,800	0.853
2004	148,673	156,000	±11,400	0.951
2005	158,982	156,000	±11,800	1.018
2006	160,114	151,000	±11,800	1.063
2007	161,662	150,000	±11,800	1.080
2008	145,875	132,000	±10,300*	1.105
2008	122,807	108,000	±9,900*	1.137

* Estimated, using equations for prior years.

Each of the three methods of gauging the completeness of reporting produces results that differ somewhat. The most rigorous evaluation—comparing actual reporting to cases that should have been reported from each State—shows a pattern of underreporting, but this result is for reporting that is three to six years old. The result of the GES comparison in Table 17 does show a steady increase in the number of cases reported, relative to the estimated national total of reportable cases. FMCSA's SSDQ evaluation program does show improvement in recent years for individual States. This improvement is consistent with the result from the comparison with the total number of reportable cases estimated using GES files. However, it must be noted that none of the methods suggests that complete and consistent reporting has been achieved. Though all indications are that crash reporting is improving. While the problem of complete crash reporting has not been resolved, there is improvement. Incomplete crash reporting affects the accuracy of the Crash Indicator BASIC that is derived from reported crashes. It is a strength of the SMS methodology that multiple indicators are available to identify motor carriers for interventions.

Reporting latency also affects the completeness of the reports available to calculate the Crash Indicator BASIC. Latency is defined as the number days between the crash event and the date the record is uploaded to the MCMIS Crash file. Records are required to be reported within 90 days. Table 18 shows the percentage of cases reported by the number of days elapsed between the crash date and the date the record was first uploaded to the MCMIS Crash file for the four original test States. The table shows that the rates of reporting vary significantly between the States. Colorado reported more than 98 percent within the 90-day requirement, but Georgia reported only about two-thirds, Missouri 88.9 percent, and New Jersey only about 83.3 percent. Within 120 days, Colorado and Missouri had reported almost all the cases they were ultimately to report, while Georgia still was missing almost 25 percent and New Jersey 8.8 percent. Even after 180 days, Georgia still had almost 20 percent of the records still to report. It is important to note that these rates are just of the cases ultimately reported, not of the records that *should have been* reported.

Table 18. Reporting Latency for Test States, 2008 Crashes

Latency in Days	Colorado	Georgia	Missouri	New Jersey
30	62.5%	2.8%	43.0%	0.4%
60	96.2%	30.9%	60.3%	47.0%
90	98.4%	66.6%	88.9%	83.3%
120	99.1%	76.5%	96.9%	91.2%
150	99.4%	77.9%	98.7%	94.8%
180	99.7%	78.8%	98.9%	96.2%

For three out of the four States, reporting latency is not a major problem. Even though only one substantially met the 90 day reporting period, three of the States reported at least 90 percent of all the cases they were to report, and at least 95 percent within 150 days of the crash. Reporting latency had a significant effect for the fourth State however, with only 78.8 percent of the eventual cases reported within 180 days. If one assumes a 90 percent reporting rate, the observed 78.8 percent latency means that only about 71 percent (0.788 times 0.9 equals 0.709) of crashes that should have been reported would have been available for the Crash Indicator BASIC after 180 days.

Reporting "timeliness," or latency is also rated by the SSDQ program. The current ratings (as of December 2010) of the four States is consistent with the analysis above. Three of the States are rated "good" and one is rated "poor." Overall, 35 States are rated "good" on timeliness of reporting, meaning that at least 90 percent of reportable cases are reported within the 90 day reporting period. Twelve are rated "fair," (65-90 percent reported within 90 days), and four are rated "poor."

The final source of data for the BASICs is the MCMIS Inspection file. The file includes the results of inspections of vehicles, drivers, and carriers. Inspections may be conducted during roadside checks, traffic enforcement stops, and carrier reviews. Data from the inspection file are used to determine all the BASICs other than the Crash Indicator.

Unlike the MCMIS Crash file, there really is no independent source of information to check the Inspection file for completeness and accuracy. In the case of the Crash file, the original State

files can be used to determine if a State is reporting all records that qualify for reporting. There is no alternative source of information on violations. Therefore, it is not possible to determine the completeness of the Inspection file. It is assumed to be complete; for practical purposes, it is assumed that there is no significant underreporting or over reporting of violations.

Results of inspections are required to be uploaded to the MCMIS Inspection file within 21 days of the inspection. Reporting latency does not appear to be an issue for the four original test States. All the States had reported more than 95 percent of inspections within 21 days. After 28 days, two of the four had reported more than 99 percent and the other two had reported more than 97 percent. Reporting was effectively 99 percent for all four States after 42 days. (See Table 19.)

Table 19. Reporting Latency for Test States, 2008 Inspections

Reporting Latency (Days)	Colorado	Georgia	Missouri	New Jersey
7	69.7%	93.7%	86.7%	77.6%
14	74.8%	97.1%	95.6%	90.3%
21	96.1%	98.5%	98.3%	95.4%
28	97.8%	99.1%	99.3%	97.5%
35	98.3%	99.4%	99.7%	98.5%
42	98.9%	99.6%	99.9%	99.0%

To serve as an optimal monitor of the safety performance of carriers, inspections would be carried out on a random basis. That way, all carriers operating on the roads would have an equal chance of being selected for inspection, and the results would be a true measure of compliance with FMCSR. However, inspections are triggered by a variety of events, including traffic enforcement stops, post-crash inspections, size and weight inspections, and local enforcement inspections. About 30 percent of the inspections are related to one of the triggers mentioned. In some States, (e.g., Michigan), probable cause is needed to stop a vehicle for an inspection. Roadside inspections are likely largely nonrandom, unless there is a specific program implemented at the State level to do random stops. Nonrandom stops tend to bias the data toward vehicles that in some way trigger a stop, such as vehicles that are speeding or have defects that are readily visible. Violations on vehicles where there is not some readily observed defect or traffic violation, (which could serve as probable cause for the stop) would be missed. It is unknown whether serious bias is introduced, without a rigorous study including random stops and inspections.

4.2 DESIGN ELEMENTS

The SMS includes a number of fundamental elements that are used in the calculation and ranking of the BASICs scores. These elements include normalization of counts of violations or crashes, the use of peer groups to rank the scores, weighting by time and by severity, and rules to determine whether the data are sufficient to calculate a BASIC score. In this section, each of the elements is discussed, to expose the rationale for the element and to consider the basic appropriateness of the element. In addition, the effect of the data sufficiency requirements in terms of the number of carriers scored is examined.

Normalization of BASICs to reflect exposure—the BASICs are "normalized" to reflect exposure. "Normalization" simply means that the metrics are expressed as a count of the relevant event per some unit of exposure. The measures of exposure used are the number of relevant inspections or the number of power units, depending on the metric. The purpose of this "normalization" is to account for the fact that carriers vary in size and operations. A very large carrier might have 10,000 power units and hundreds of inspections, while a small carrier might operate only one or two power units and have only a handful of inspections. Normalization is the standard way to account for variation in exposure to risk.

Two measures are used for normalization, depending on the BASIC. The count of power units is used for the Unsafe Driving (BASIC 1), Controlled Substances (BASIC 4), and the Crash Indicator. The number of relevant inspections is used for the other BASICs: Fatigued Driving (BASIC 2), Driver Fitness (BASIC 3), Vehicle Maintenance (BASIC 5), and Improper Loading/Cargo Securement (BASIC 6).

The count of power units used for exposure includes power units owned, term-leased, or trip-leased—in other words, the total number of power units operated by the authority of the carrier, which is appropriate. In addition, the number of power units used is the average number over the previous 18 months. For any given carrier, the number can vary in response to business conditions, such that a snapshot at a particular time can be misleading a few months later. Averaging power units at three points in the prior 18 months is a reasonable way to account for the variation over time.

The use of inspections to normalize counts of violations is also reasonable, though subject to limitations related to reporting latency and nonrandom inspections. Reporting latency seems minimal, as shown in Table 19 above. The effect of nonrandom inspections is more related to whether enough inspections are carried out for a particular carrier to meet the data sufficiency requirements, discussed below. Some carriers might be overlooked and thus be less likely to have enough inspections to meet the data sufficiency requirements, while others would have more inspections than expected if the inspections were purely random. However, there is no reason to believe that would affect the results of the inspections themselves.

Weighting by violation and by crash severity—violations and crashes are weighted by their severity and association with crash risk. The decision to weight violations and crashes is eminently reasonable. Clearly fatal crashes are much worse than nonfatal crashes and therefore should provoke a more rigorous response. And it is also clear, just based on experience, that not all of the hundreds of FMCSR bear the same relationship to crash risk, and that the focus should be on those that contribute more to crash risk and to crash severity.

Whether the weights used in the calculation of the BASICs scores are appropriate is not known, however. With respect to crash severity, fatal and injury crashes are assigned a weight of two, and towaway crashes a weight of one. If hazardous materials were involved in the crash and released, one is added to the severity score. In the document describing the SMS, no rationale or justification for the weights are given.[6] One source for the weights could be to use crash costs associated with different crash severities. For example, Zaloshnja and Miller have developed comprehensive costs for different truck and bus crash types and severities.[7] These would provide a substitute based on analysis for the arbitrary weights currently used.

Choice of weights for violations appears to rest on a set of analyses of crash occurrence, consequences, and effectiveness, all subject to subject matter expert review.[6] These analyses have not been reviewed for this evaluation, and so no judgment is made here as to their reliability. However, based on our own experience with the data available for such an analysis, this is an area that would benefit from continuing review and improvement. The weights themselves seem reasonable, but given the sheer number of violations and the difficulty of linking individual violations to crash risk, this is clearly an area for continued development as the SMS goes forward.

Time weights for safety events—safety events (crashes or violations) are weighted by time, i.e., recent crashes or violations are weighted more than crashes or violations more distant in time. The time window for crashes and violations is the previous 24 months. For all metrics, events more than 24 months old are excluded from consideration. Within the-24 month window, three time blocks are defined. Specific safety events within the previous 6 months are given a weight of 3; those occurring between 6 and 12 months prior are assigned a weight of 2; and those between 12 and 24 months receive a weight of 1.

It is reasonable on its face to assign greater weight to recent safety events than older ones, since the purpose of the CSA 2010 process is to identify current safety issues and to move carriers toward safer behavior in the future. The choice of time periods and the weight values assigned are based on expert judgment and fairness. As the SMS is employed, it may be possible to refine the time weights based on analysis that estimates whether the association of past violations or crashes with future safety problems decreases over time.

Peer groups—BASIC percentiles are assigned within peer groups, rather than with respect to all carriers. That is, for each BASIC, scores are ranked within subsets of carriers, based on either the number of power units or the number of inspections, depending on the BASIC. The effect of using peer groups is that carriers or drivers are only compared with other carriers or drivers that are similar in terms of the method of forming the peer group.

Table 20 shows the methods used to assign peer groups for each of the metrics. The peer groups for BASICs 1, 4, and the Crash Indicator (BASIC 7) are based on the number of power units; the peer groups for BASICs 2, 3, 5, and 6 are based on the number of inspections. The relevant factor for peer groups for each BASIC is the same as the one used in normalization.

Note: At the time this analysis was performed, the peer grouping used for the Crash BASIC was based on power units. It is the researchers' understanding, however, that the FMCSA has recently changed the method of peer grouping for the Crash BASIC to a scheme based on a carrier's number of crashes. The researchers' understanding is that the change was made because the Agency believes a carrier's number of crashes is a better proxy for exposure on the road than the number of power units. Because of the timing of this change, an assessment of this is outside the scope of this study.

Table 20. Peer Groups Used in BASICs

Peer Group	Power Units (BASIC 1, 4, 7)	Relevant Inspections (BASIC 2, 3, 5, 6)
1	0–5	3–10 (BASIC 2); 5–10 (BASIC 3, 5, 6)
2	6–15	11–20
3	16–50	21–100
4	51–500	101–500
5	More Than 500	More Than 500

The rationale for using peer groups when determining percentiles is to "account for differences among carriers or drivers."[6] Motor carriers are diverse in their operational characteristics, ranging from only a single power unit to many tens of thousands, and from purely regional or dedicated contract operations to nationwide common carriers. Categorizing by the number of power units does reasonably group carriers that share at least some operational characteristics. Smaller carriers are less likely to have dedicated safety officers and the owners and operators are more likely to have to deal with the whole range of activities in running a trucking firm. In contrast, large carriers typically have safety directors who are responsible for the safe operation of the fleet. The use of peer groups also helps account for differences in exposure. Rates for small carriers are inherently more variable than for large carriers with much higher levels of exposure. By using peer groups, the SMS compares carriers with others that are similarly situated. The specific categories chosen were based on the judgment of the SMS designers. No analysis is cited in justification of the categories, though they are reasonable on their face.

Data sufficiency—data sufficiency rules are designed to ensure that the BASIC scores and percentiles are based on enough data to be reliable and to reflect the underlying pattern of violations of the carriers.[6] The rules require both a minimum number of inspections (or crashes, in the case of the Crash Indicator), and at least one violation. Table 21 shows the data sufficiency rules for each BASIC.

Table 21. Data Sufficiency Requirements for Each BASIC

BASIC	Data Requirement
BASIC 1, Unsafe Driving	3 or More Relevant Inspections; at Least One Relevant Violation
BASIC 2, Controlled Substances	At Least One Relevant Violation
BASIC 3, Fatigued Driving	3 or More Relevant Inspections; at Least One Relevant Violation
BASIC 4, Driver Fitness	5 or More Relevant Inspections; at Least One Relevant Violation
BASIC 5, Vehicle Maintenance	5 or More Relevant Inspections; at Least One Relevant Violation
BASIC 6, Improper Loading	5 or More Relevant Inspections; at Least One Relevant Violation
BASIC 7, Crash Indicator	2 or More Crash Involvements

Most carriers, in both the test States and nonparticipating States, do not meet the data sufficiency requirements to determine a BASIC score and percentile of any of the BASICs. In fact, for each of the BASICs, an overwhelming majority of carriers do not receive a BASIC score. Table 22 shows the percentage of carriers with BASIC percentiles for each of the BASICs. The percentages are shown both for test State carriers and for nonparticipating carriers (those in all the other States). The proportions with sufficient data are below 10 percent of carriers for all the

BASICs and below five percent for a majority of the BASICs. The pattern is the same for both the test States and for nonparticipating carriers. Only 11.0 percent of carriers in the test have a valid BASIC percentile for at least one BASIC.

Table 22. Percent of Carriers Meeting Data Sufficiency Requirements by BASIC, Test States and Nonparticipating Carriers

BASIC	Percent Carriers With Sufficient Data: Test States	Percent Carriers With Sufficient Data: Nonparticipating States
BASIC 1 (Unsafe Driving)	4.8%	3.8%
BASIC 2 (Controlled Substances)	6.4%	5.4%
BASIC 3 (Fatigued Driving)	2.5%	1.0%
BASIC 4 (Driver Fitness)	1.9%	0.3%
BASIC 5 (Vehicle Maintenance)	8.7%	7.8%
BASIC 6 (Improper Loading/Securement)	3.5%	2.4%
BASIC 7 (Crash Indicator)	3.4%	1.9%
At Least One BASIC	11.0%	10.4%
N	63,824	755,076

Of course, the data sufficiency requirements explain the low percentage of carriers with BASIC percentiles. The data sufficiency rules call for both a minimum number of inspections and as well as at least one violation recorded. To receive a BASIC score, a carrier must both have been inspected and have been found in violation of at least one FMCSR. That is, only carriers with violations are given a BASIC score, and then only if subjected to a minimum number of inspections. Carriers that have inspections, regardless of the number, and no violations are not included when the BASIC percentiles are determined.

The two part data sufficiency test (a minimum number of inspections and at least one violation) reflects the fact that the data sufficiency rules really incorporate two different ideas. The requirement for a minimum number of inspections is designed to take only carriers that have been looked at enough that the inspection results can be considered reliable.[(6)] The second idea is to focus only on carriers with violations, and to exclude carriers with no violations from the percentile rankings.

Table 23 shows the results of 15 months of inspections (January 2008–March 2009). The proportion of carriers with inspections that resulted in no violations uncovered is modest. In the test States, only 8.7 percent of inspected carriers had one or more inspection with no violations. The proportion was similar (9.2 percent) for carriers in the non-test States. But most of those were either one or two inspections. Of test State carriers with no violations, only 0.7 percent had three or four inspections, and only 0.2 percent had five or more inspections. So while there may be some concern that carriers with perfect records are excluded from the result, there is little substantive impact.

Table 23. Percent Distribution of Carriers by Number of Violations and of Inspections, Test State and Nonparticipating Carriers

Number of Violations	Number of Inspections	Test States	Nonparticipating
No Violations	1 or 2	7.8%	8.4%
No Violations	3 or 4	0.7%	0.6%
No Violations	5 or more	0.2%	0.2%
One or More Violations	1 or 2	44.1%	46.9%
One or More Violations	3 or 4	17.3%	15.9%
One or More Violations	5 or more	30.0%	28.0%
Total		**100.0%**	**100.0%**
N		**19,352**	**369,702**

The consequence of excluding carriers with only a small number of inspections is more substantial. Over 44 percent of test State carriers had a violation but only one or two inspections, and thus did not meet the data sufficiency requirements. Another 17.3 percent of inspected carriers had only three or four inspections and so would not meet the data sufficiency requirements for BASICs 4, 5, and 6. Only 30.0 percent of inspected carriers had five or more inspections.

And note that Table 23 includes only carriers with inspections. About 30.3 percent of test State carriers had any inspection over the 15 month period, meaning that about 70 percent of the carriers were not touched at all. Thus, about 70 percent of the low rates of data sufficiency are related to the fact that most registered carriers in the test States received no inspections. The reason for this is unknown, though doubtless some of the carriers operate at such a low level with so few vehicles that none of their vehicles or drivers were inspected by chance.

Table 24 shows that rates of data sufficiency are strongly associated with carrier size. The rates for small carriers are quite low, with only 0.2 percent to 3.6 percent of carriers with five or fewer power units meeting the data sufficiency rules, depending upon the BASIC. Rates are much higher for larger carriers. More than 45 percent of carriers with 16–50 power units meet the data sufficiency rules for at least one BASIC. Almost 80 percent of the largest carriers (500 or more power units) meet the requirements for at least one BASIC. Thus the data sufficiency rules tend to catch the carriers with the most power units and that operate most trucks.

Table 24. Data Sufficiency Rates, by BASIC and Carrier Size, Test States

Carrier Size	BASIC 1	BASIC 2	BASIC 3	BASIC 4	BASIC 5	BASIC 6	BASIC 7	Any BASIC
0–5 PU	1.1%	3.1%	0.2%	0.1%	3.6%	0.4%	0.3%	5.7%
6–15 PU	9.9%	11.6%	2.2%	0.4%	23.4%	5.5%	4.0%	28.3%
16–50 PU	27.4%	25.1%	6.1%	1.3%	45.1%	17.2%	17.7%	50.2%
51–500 PU	47.9%	40.0%	22.9%	4.3%	59.1%	37.2%	46.7%	65.7%
500+ PU	71.4%	55.1%	63.3%	20.4%	79.6%	67.3%	77.6%	83.7%
Missing Data	19.1%	19.1%	19.1%	19.1%	19.2%	19.1%	19.1%	19.2%
Total	**4.8%**	**6.4%**	**2.5%**	**1.9%**	**8.7%**	**3.5%**	**3.4%**	**11.0%**

The data sufficiency rules are a reasonable effort to address the problem of small sample sizes. Only one or two inspections of a carrier's vehicles or drivers is on its face too few to provide a reliable reading on its operations. However, the rules should be understood in the context of the number of inspections that are done and the probability that a carrier would be subject to an inspection. About 60 percent of carriers received no inspections in the 15 months of data reviewed. Probably most of those carriers were operating at such a low level, with so few trucks, that they escaped inspection. Of the carriers that were inspected, more than half had only one or two inspections, and so failed the data sufficiency rules on that basis. Almost 45 percent of the inspected carriers had a violation, but only one or two inspections.

However, it is noteworthy that the net impact of the rules excludes more small carriers. Only 5.7 percent of carriers with fewer than six power units meet the data sufficiency rules to be assigned a percentile on any BASIC. In contrast, more than 50 percent of carriers with 16–50 power units met the sufficiency requirements on at least one BASIC, as did almost 84 percent of carriers in the largest category, with more than 500 power units. It should be kept in mind, however, that most trucks are operated by large carriers. Carriers with 100 or more power units operate about half of all trucks. While small carriers are more numerous, large carriers operate a disproportionate share of the trucks on the road.

5. THE SMS BASICS AND IDENTIFICATION OF UNSAFE CARRIERS

The goal of this section is to determine to what extent the BASIC percentiles calculated under the new SMS identify unsafe carriers. Since the ultimate goal of the FMCSA is to reduce fatalities and injuries resulting from CMV crashes, analysis of the relationships between crash rates and BASIC percentiles can help determine if the SMS identifies high-risk carriers. Therefore, crash rates are calculated for carriers that exceed the BASIC thresholds, and are compared to the crash rates of carriers that do not exceed BASIC thresholds. If the SMS is successful at identifying high risk carriers, then the crash rates for carriers exceeding BASIC thresholds should be significantly higher than those for carriers not exceeding BASIC thresholds.

Another comparison of interest is whether the SMS does a better job of identifying unsafe carriers than the current SafeStat method. Unlike SafeStat, however, which scores motor carriers in four SEAs and ranks carriers according to a weighted average of unacceptable SEAs, the SMS is designed to identify specific types of unsafe behaviors by measuring carriers and drivers in the seven BASICs. One of the goals of the CSA 2010 test is to touch more carriers that exceed BASIC thresholds through specific interventions such as warning letters, onsite and offsite investigations, and CSPs. Therefore, crash rates and numbers of carriers identified by SafeStat and the new SMS are compared.

The BASIC percentiles that are calculated under the new measurement system use data from the MCMIS data files. The FMCSA depends on the individual States to upload data on crashes and violations to the MCMIS database in a timely manner. Based on evaluations performed by the researchers, it is known that there is variability in reporting qualifying crashes to the MCMIS Crash file among the States. Since calculation of the BASIC percentiles depends on these data, effects of late reporting to the MCMIS Crash file are investigated.

In addition to comparing crash rates between groups of carriers exceeding BASIC thresholds and carriers not exceeding BASIC thresholds, relationships between crash rates and BASIC percentiles are explored. If there are positive associations between crash rates and BASIC percentiles, then as BASIC percentiles increase, crash rates should increase. First, for each BASIC, scatter plots are created showing the association between crash rates and BASIC percentiles. Then a statistical model is developed that predicts crash rates from all six of the BASIC percentiles, excluding the Crash Indicator. The model can be used to predict a level of safety for carriers based on the BASIC percentiles exceeded.

To assess the magnitudes of crash rates for carriers identified under the new SMS, crash rates were calculated for carriers classified as *nonparticipating* by the CSA 2010 Operational Model Test from February 2008 through July 2009. For the entire 18 months, these carriers were classified as nonparticipating, meaning they were not included in the CSA 2010 test and should not have received CSA 2010 interventions. On the other hand, they may have received CRs under the current SafeStat system used for identifying high-risk carriers. The reason for choosing the nonparticipating carriers is that they represent a large percentage of all carriers in the United States and are involved in sufficient numbers of crashes to make sample sizes large enough for analysis. To minimize any effects of late reporting to the MCMIS Crash file, the file is dated

June 2010, so crashes occurred at least 11 months prior to the date of the file, and should have been uploaded by the States.

Carriers in the four original States in the Operational Model Test plus two States added as test-only States during the time period that rates are calculated (Minnesota and Montana), are excluded from analysis. In addition, the analysis is restricted to the approximate 500,000 carriers with recent activity. For data quality purposes, certain carriers with outlying data that can have a large impact on group crash rates are excluded. Typically, these carriers have a large number of power units, but no reported crashes due to their type of operation (e.g., MCMIS data shows that a carrier that leases trucks has approximately 30,000 power units, but no crashes. Such a large number of power units with no reported crashes has a large influence on crash rates).

5.1 DOES THE SMS IDENTIFY UNSAFE CARRIERS?

Table 24 shows 18-month crash rates for 473,847 carriers with recent activity according to their BASIC percentile scores determined during February 2008. For each BASIC, carriers are classified as to whether they exceeded that particular BASIC threshold. Note that a carrier may exceed more than one BASIC threshold and may be counted several times among the various carrier groups. For comparison, crash rates are also calculated for carriers that exceeded any BASIC threshold, and those that exceeded no BASIC thresholds. Since those two categories are mutually exclusive, numbers will sum to totals. The last column of the table gives the ratio of the crash rate to the crash rate for carriers that exceeded no BASIC thresholds.

The crash rates for carriers exceeding BASIC thresholds are significantly higher than for carriers exceeding no BASIC thresholds. The crash rate for carriers exceeding the Unsafe Driving threshold is 7.44, which is greater than the crash rate for carriers exceeding the Crash Indicator BASIC, and is 3.56 times greater than the rate for carriers exceeding no BASIC thresholds. Crash rates for carriers exceeding the Fatigued Driving, Controlled Substance and Alcohol, and Vehicle Maintenance BASICs are also high relative to the 2.09 crash rate for carriers exceeding no BASIC thresholds. Also note that the numbers of carriers exceeding the Vehicle Maintenance, Fatigued Driving, Improper Loading/Cargo Securement, and Unsafe Driving thresholds are relatively large. Exceeding the Controlled Substance and Alcohol BASIC is relatively rare. The total number of carriers exceeding any BASIC threshold is 44,881 (9.5 percent) of all carriers with recent activity. Note that this is close to the estimated 9.9 percent expected touches by CSA based on data from the original four test States given in Table 14. Using the crash rate as a measure of risk, it appears that the SMS tends to identify unsafe carriers.

Table 25. SMS Crash Rates Calculated During the 18-Month Followup Period (February 2008–July 2009), Based on SMS Classification February 2008 for Nonparticipating Carriers

BASIC Threshold Exceeded	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
Unsafe Driving	9,245	33,532	450,874	7.44	3.56
Fatigued Driving	17,959	15,525	248,862	6.24	2.99
Driver Fitness	3,981	11,539	379,009	3.04	1.46
Controlled Substance and Alcohol	1,013	6,860	104,799	6.55	3.14
Vehicle Maintenance	18,280	13,643	278,198	4.90	2.34
Improper Loading/Cargo Securement	9,409	16,747	421,670	3.97	1.90
Crash Indicator	5,077	33,946	463,766	7.32	3.51
Exceeded Any BASIC	44,881	63,452	1,284,475	4.94	2.37
Exceeded No BASICs	428,966	45,029	2,157,939	2.09	1.00
All Carriers	473,847	108,481	3,442,414	3.15	1.51

5.2 COMPARISON BETWEEN THE SMS AND SAFESTAT CRASH RATES

For comparative purposes, crash rates were calculated for the same 473,847 carriers based on SafeStat classification. Table 25 shows 18-month crash rates from February 2008 to July 2009 for carriers according to their SafeStat status determined during February 2008. SafeStat uses 30 months of data to compute various measures and indicators from the source data that are combined into the four SEAs. Based on the February 2008 ranking, carriers are grouped into three safety risk categories: SafeStat A/B, SafeStat C, and SafeStat A/B/C. For comparison, crash rates are also calculated for SafeStat Not Identified carriers which are those not classified as A, B, or C. The last column in Table 25 gives the ratio of the crash rate to the Not Identified group.

Of the 473,847 carriers that remained nonparticipating during the 18-month followup period, 5,402 are in the A/B category, 3,389 are in the C category, 8,791 are in the A/B/C category, and 465,056 are Not Identified. Based on the results in Table 25, the crash rate for the A/B carrier group is 2.30 times greater during the followup period than the carriers not identified. The crash rate for the C carrier group is 1.64 times greater than the carriers not identified.

Table 26. SafeStat Crash Rates Calculated During the 18-Month Followup Period (February 2008–July 2009), Based on SafeStat Classification February 2008

Carrier Group	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
SafeStat A/B	5,402	6,605	95,159	6.94	2.30
SafeStat C	3,389	2,764	55,900	4.94	1.64
SafeStat A/B/C	8,791	9,369	151,059	6.20	2.06
SafeStat Not Identified	465,056	99,112	3,291,355	3.01	1.00
All Carriers	**473,847**	**108,481**	**3,442,414**	**3.15**	**1.05**

The SMS identifies many more carriers for intervention than does SafeStat. The number of SafeStat A/B/C identified carriers is 8,791, while the number of carriers exceeding any SMS BASIC threshold is 44,881. In addition, several of the BASIC crash rates are comparable to the SafeStat A/B crash rate. In particular, the crash rates for carriers exceeding the Unsafe Driving and the Controlled Substance and Alcohol BASIC thresholds are 7.44 and 6.55, respectively. The crash rate for carriers exceeding the Fatigued Driving crash rate is 6.24. The crash rate for the 428,966 carriers that exceeded no BASIC thresholds is 2.09 which is less than the 3.01 crash rate for carriers not identified by SafeStat. Therefore, the ratios to not identified measures in the last column of Table 24 tend to be larger than those shown in Table 25.

5.3 THE EFFECT OF CRs ON GROUP CRASH RATES

Carriers with recent activity classified as nonparticipating were chosen to evaluate the SMS because they represent almost the entire population of carriers in the United States, and they did not participate in the CSA 2010 Operational Model Test. Therefore, they provide an adequate sample of carriers for analysis, and they did not receive interventions under CSA 2010. However, they may have received CRs under the current method used to identify high-risk carriers.

To remove the possible effects on crash rates of CRs during the 18-month followup period, crash rates were calculated for carriers that did not receive CRs during the followup period. Table 27 shows SMS crash rates for the carriers that did not receive CRs during the followup period. Although not by a great amount, the crash rates in Table 27 are less than the crash rates in Table 24 for each BASIC category. However, the last column of the table shows that crash rates for carriers exceeding BASIC thresholds are several times higher than for carriers exceeding no BASIC thresholds. It appears that the SMS still tends to identify high-risk carriers even after removing the effects of CRs. The result may not be surprising because the number of CRs over the period is 473,847 minus 457,740 equals 16,107 which is a small percentage of total carriers.

Table 27. SMS Crash Rates Calculated for Carriers Without CRs During the 18-Month Followup Period (February 2008–July 2009), Based on SMS Classification February 2008

BASIC Threshold Exceeded	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
Unsafe Driving	6,968	24,008	345,268	6.95	3.50
Fatigued Driving	14,009	8,637	169,206	5.10	2.57
Driver Fitness	2,838	8,117	325,546	2.49	1.26
Controlled Substance and Alcohol	724	4,598	77,336	5.95	3.00
Vehicle Maintenance	14,550	8,838	216,767	4.08	2.05
Improper Loading/Cargo Securement	7,650	12,495	358,707	3.48	1.76
Crash Indicator	3,363	24,991	360,322	6.94	3.49
Exceeded Any BASIC	36,786	47,974	1,076,153	4.46	2.25
Exceeded No BASICs	420,954	41,001	2,065,964	1.98	1.00
All Carriers	**457,740**	**88,975**	**3,142,117**	**2.83**	**1.43**

In a similar fashion, crash rates were calculated for SafeStat group carriers that did not receive CRs during the followup period. Again, the crash rates are lower in every category when compared to the rates for carriers including those with CRs. However, the ratio to not identified column is very similar to the one that includes carriers with CRs. It appears that removing carriers without CRs does not substantially change the results presented.

Table 28. SafeStat Crash Rates Calculated for Carriers Without CRs During the 18-Month Followup Period (February 2008–July 2009), Based on SafeStat Classification February 2008

Carrier Group	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
SafeStat A/B	2,499	2,400	38,741	6.19	2.23
SafeStat C	2,250	1,452	34,926	4.16	1.50
SafeStat A/B/C	4,749	3,852	73,667	5.23	1.88
SafeStat Not Identified	452,991	85,123	3,068,450	2.77	1.00
All Carriers	**457,740**	**88,975**	**3,142,117**	**2.83**	**1.02**

5.4 RATE RATIOS COMPARING CARRIERS EXCEEDING THRESHOLDS TO CARRIERS NOT EXCEEDING THRESHOLDS

A slightly different way of assessing SMS effectiveness is to compare the crash rate for carriers exceeding SMS thresholds to carriers not exceeding thresholds for each BASIC. Table 29 shows crash rates for carriers exceeding and not exceeding a particular threshold, and in the rightmost column the ratio of the two rates. The rate ratios for the Unsafe Driving BASIC and the Crash Indicator BASIC are similar and suggest that crash rates are about 3 times larger for carriers exceeding these two thresholds. The rate ratios for the Fatigued Driving and the Substance/Alcohol BASICs are also similar indicating that crash rates are about twice as large for carriers that exceed these thresholds. The Vehicle Maintenance and Loading/Cargo BASICs have rate ratios less than two. The rate ratio for the Driver Fitness BASIC is close to one, suggesting no difference in crash rates for carriers exceeding this BASIC compared to carriers not exceeding this BASIC. For carriers exceeding any threshold (at least one) the crash rate is about 2.37 times higher than for carriers exceeding no thresholds.

Table 29. Crash Rates and Rate Ratios for Carriers Exceeding Thresholds and Carriers not Exceeding Thresholds Calculated During the 18-Month Followup Period (February 2008–July 2009), Based on SMS Classification February 2008 for Nonparticipating Carriers

Basic	Threshold Exceeded Carriers	Threshold Exceeded Crashes	Threshold Exceeded Power Units	Threshold Exceeded Crash Rate Per 100 PU	Threshold Not Exceeded Carriers	Threshold Not Exceeded Crashes	Threshold Not Exceeded Power Units	Threshold Not Exceeded Crash Rate Per 100 PU	Rate Ratio
Unsafe Driving	9,245	33,532	450,874	7.44	464,602	74,949	2,991,540	2.51	2.97
Fatigued Driving	17,959	15,525	248,862	6.24	455,888	92,956	3,193,552	2.91	2.14
Driver Fitness	3,981	11,539	379,009	3.04	469,866	96,942	3,063,405	3.16	0.96
Substance/Alcohol	1,013	6,860	104,799	6.55	472,834	101,621	3,337,615	3.04	2.15
Vehicle Maintenance	18,280	13,643	278,198	4.90	455,567	94,838	3,164,215	3.00	1.64
Loading/Cargo	9,409	16,747	421,670	3.97	464,438	91,734	3,020,744	3.04	1.31
Crash Indicator	5,077	33,946	463,766	7.32	468,770	74,535	2,978,648	2.50	2.93
Any BASIC	44,881	63,452	1,284,475	4.94	428,966	45,029	2,157,939	2.09	2.37

5.5 UNDERREPORTING AND LATE DATA REPORTED TO THE MCMIS CRASH FILE

The FMCSA depends on the States to upload data on crashes and roadside inspections to the MCMIS database. Calculation of the SMS BASICs depends on this information. All reportable crash involvements for a calendar year are required to be transmitted to the MCMIS Crash file within 90 days of the date of the crash. The requirement for inspection data is within 21 days. The focus here is on late reporting to the crash file since it generally takes longer to receive the crash data. In addition, calculation of the Crash Indicator BASIC depends on data reported to the MCMIS Crash file.

In 2003, researchers began evaluating data reported by the States to the MCMIS Crash file.[8,9,10,11] Although reporting rates tend to vary from State to State, the studies have shown some consistent trends in underreporting. For example, fatal involvements are much more likely to be reported than those involving injury or those in which vehicles are towed due to disabling damage. Crashes involving tractor combinations are more likely to be reported than those involving straight trucks. Reporting to the crash file can also depend on county size or the law enforcement agency responsible for reporting the crash.

Figure 3 shows total crashes reported to the MCMIS Crash file over two consecutive 12-month periods for carriers with USDOT numbers. The MCMIS file is dated June 2010. The dashed line represents crashes that occurred from July 2009 through June 2010. The solid line represents crashes that occurred one year earlier from July 2008 through June 2009. The solid line shows about 8,000 crashes per month, with a spike to more than 10,000 in December 2008, and then a decline to about 7,000 through June 2009. The important note is that the solid line is fairly consistent over time, except for a slight decline that may be due to reduced crashes, indicating that most of the crashes in the earlier period have been uploaded to the MCMIS Crash file by June 2009. However, the dashed line shows a steep decline in reported crashes over the 3 months prior to June 2010. There is a 90-day grace period for individual States to upload crashes to the file, but calculation of the CSA 2010 Crash Indicator BASIC percentiles depend on a 24-month window of crashes, which includes the 3 most recent months of data recorded in the file. As long as systematic differences among entities being compared remain constant over time, the time lag may not pose any major difficulties. Although reporting by the States is improving over time, a bigger concern than late reporting is likely underreporting.[10,11]

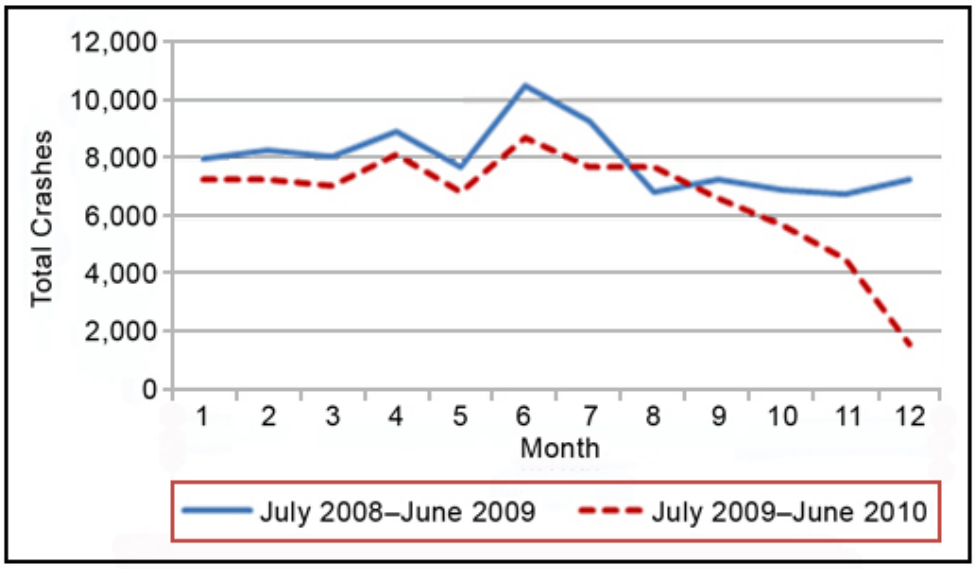

Figure 3. Crashes Reported to the MCMIS Crash File by Month as of June 2010

Table 30 shows crashes reported to the MCMIS Crash file by State in March 2010, as a percentage of the average number of crashes over the 3-month period from April 2009 through June 2009. The MCMIS Crash file being used here is dated June 2010, so March 2010 is outside the 90-day grace period in which States should have uploaded crashes to the file. The 3-month average of crashes between April and June 2009 represents the typical number of crashes that a State reports in a month, and data reported over that time period should be fairly complete, except for any underreporting, since June 2009 is 1 year prior to the date of the file. It should be pointed out that Table 30 is only a *snapshot* of a point in time, and results can vary by month.

Carriers in States that do not report crashes in a timely manner are less likely to be identified by the Crash Indicator BASIC. For example, based on historical data, Maryland typically reports about 114 crashes per month to the crash file. However, by June 2010, Maryland reported only one crash that occurred in March of the same year. Using the same criteria, other States such as Arkansas and Georgia have reporting percentages less than 50 percent.

Table 30. MCMIS-Reported Crashes in March 2010 by State as a Percentage of 3-Month Average Reported Crashes (April 2009–June 2009), From the June 2010 MCMIS Crash File

State	April 2009	May 2009	June 2009	3-Month Average	Reported March 2010	Percent Reported
AK	8	3	7	6.0	5	83.3%
AL	141	152	126	139.7	179	128.2%
AR	101	117	89	102.3	43	42.0%
AZ	145	130	150	141.7	141	99.5%
CA	375	348	361	361.3	395	109.3%
CO	104	89	105	99.3	120	120.8%
CT	38	52	62	50.7	47	92.8%
DE	33	24	24	27.0	15	55.6%
FL	294	168	245	235.7	329	139.6%
GA	290	325	302	305.7	151	49.4%
HI	4	2	4	3.3	4	120.0%
IA	111	109	95	105.0	114	108.6%
ID	33	27	36	32.0	45	140.6%
IL	338	352	407	365.7	331	90.5%
IN	246	235	266	249.0	240	96.4%
KS	97	83	119	99.7	83	83.3%
KY	191	183	177	183.7	182	99.1%
LA	183	165	170	172.7	221	128.0%
MA	109	136	114	119.7	94	78.6%
MD	94	129	120	114.3	1	0.9%
ME	21	28	42	30.3	27	89.0%
MI	216	211	212	213.0	182	85.4%
MN	122	114	136	124.0	90	72.6%
MO	211	251	239	233.7	256	109.6%
MS	96	79	64	79.7	65	81.6%
MT	26	29	23	26.0	21	80.8%
NC	243	220	250	237.7	222	93.4%
ND	6	20	16	14.0	21	150.0%
NE	71	64	80	71.7	57	79.5%
NH	20	22	23	21.7	23	106.2%
NJ	301	279	341	307.0	346	112.7%
NM	32	32	32	32.0	40	125.0%
NV	46	30	38	38.0	26	68.4%
NY	174	236	196	202.0	174	86.1%
OH	211	231	286	242.7	299	123.2%
OK	131	136	149	138.7	164	118.3%
OR	63	58	65	62.0	59	95.2%

State	April 2009	May 2009	June 2009	3-Month Average	Reported March 2010	Percent Reported
PA	360	367	375	367.3	370	100.7%
RI	22	11	15	16.0	14	87.5%
SC	109	130	155	131.3	163	124.1%
SD	18	19	15	17.3	27	155.8%
TN	215	224	216	218.3	171	78.3%
TX	621	593	655	623.0	427	68.5%
UT	88	74	94	85.3	71	83.2%
VA	224	220	252	232.0	210	90.5%
VT	9	10	10	9.7	5	51.7%
WA	68	81	75	74.7	74	99.1%
WI	109	126	132	122.3	112	91.6%
WV	58	53	58	56.3	53	94.1%
WY	63	15	28	35.3	86	243.4%

To determine what impact late data has on the calculation of crash rates used to assess the SMS and SafeStat, the analyses shown previously were repeated using data from the latest months of the MCMIS Crash file. That is, the data are taken from the period January 2009 through June 2010 and include the last three months shown by the dashed line in Figure 3. Therefore, crash rates are calculated over an 18-month followup period beginning in January 2009, at which time BASIC percentiles and SafeStat status are recorded, and ending with June 2010, the date of the file.

The same types of carriers are identified for analysis—carriers with recent activity that were nonparticipating for the 18-month followup period, from 44 States and the District of Columbia. This provides the sample sizes necessary to calculate stable rates. However, some of these carriers are not the same carriers used in the previous analysis because the data are from a different time period. The total number of identified nonparticipating carriers using these data is 447,371, compared to 473,847 in the previous analysis.

Table 31 shows SMS crash rates calculated during the 18-month followup period using data based on BASIC percentiles recorded January 2009. Since fewer crashes have been uploaded to the MCMIS Crash file using the latest data, crash rates are smaller for each category when compared to Table 24. However, the ratios to not identified measures, which represent relative risks and not absolute risks, are of similar magnitude to those shown in Table 24. Therefore, in this aggregate analysis of current data, it appears that the SMS tends to identify high risk carriers.

Table 31. SMS Crash Rates Calculated During the 18-Month Followup Period (January 2009–June 2010), Based on SMS Classification January 2009

Carrier Group	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
Unsafe Driving	8,914	27,823	446,296	6.23	3.74
Fatigued Driving	17,741	13,303	243,500	5.46	3.27
Driver Fitness	3,793	10,154	404,292	2.51	1.50
Controlled Substance and Alcohol	878	4,938	94,486	5.23	3.13
Vehicle Maintenance	18,613	13,644	310,377	4.40	2.63
Improper Loading/Cargo Securement	9,388	13,359	437,012	3.06	1.83
Crash Indicator	4,775	28,556	467,532	6.11	3.66
Exceeded Any BASIC	44,680	53,122	1,305,240	4.07	2.44
Exceeded No BASICs	402,691	34,537	2,069,186	1.67	1.00
All Carriers	447,371	87,659	3,374,426	2.60	1.56

For the same 447,371 active and nonparticipating carriers identified above, Table 32 shows the repeated analysis for SafeStat identified carriers. As for the SMS results, crash rates are smaller than those in Table 25 for each category due to the smaller number of reported crashes. However, there are not substantial differences in the last columns of Table 25 and Table 31.

Table 32. SafeStat Crash Rates Calculated During the 18-Month Followup Period (January 2009–June 2010), Based on SafeStat Classification January 2009

Carrier Group	Carriers	Crashes	Power Units	Crash Rate Per 100 PU	Ratio to Not Identified
SafeStat A/B	5,483	5,064	86,065	5.88	2.37
SafeStat C	3,648	2,306	59,114	3.90	1.57
SafeStat A/B/C	9,131	7,370	145,179	5.08	2.04
SafeStat Not Identified	438,240	80,289	3,229,247	2.49	1.00
All Carriers	447,371	87,659	3,374,426	2.60	1.04

These aggregate analyses tend to demonstrate that the BASIC measures can be used to identify high risk carriers. At the individual carrier level, however, some high risk carriers may not be identified. For example, carriers in States that are late in reporting crashes to the MCMIS Crash file might not be identified by the Crash Indicator BASIC. This would tend to penalize carriers in States that do a good job of reporting.

5.6 SMS AND SAFESTAT CLASSIFICATION

The following analysis is an attempt to compare carriers in the Operational Model Test based on whether they were flagged as SafeStat A/B carriers or whether they exceeded any BASIC thresholds. Table 33 shows a 2x2 cross tabulation of 81,067 carriers that were test or control group carriers in the original four States at any time during the 29 months of the Operational Model Test. The carriers are categorized as to whether they exceeded at least one BASIC

threshold or whether they were classified as a SafeStat A/B carrier at least once during the time that they were test or control carriers.

All of the 81,067 carriers in Table 33 may not have participated in the Operational Model Test for the entire 29 months. Some may have been removed or excluded from the test at some time, but they are only evaluated over the period in which they were test or control carriers. The table shows that 121 carriers were classified as SafeStat A/B carriers at least once over the time period, but exceeded no BASIC thresholds over the same time frame. On the other hand, 9,521 carriers were identified by the new SMS as exceeding at least one BASIC threshold over the time frame, but were not classified as SafeStat A/B over the same time period. Therefore, the new SMS identifies many carriers for intervention that are not identified as SafeStat A/B carriers.

Table 33. Test and Control Carriers Categorized by SafeStat and BASIC Classifications Over the 29 Months of the Operational Model Test (Original Four States: CO, GA, MO, NJ)

Exceeded Any BASIC	SafeState A/B At Least Once	SafeState A/B None	Total
At Least Once	1,776	9,521	11,297
None	121	69,649	69,770
Total	1,897	79,170	81,067

5.7 ASSOCIATIONS BETWEEN BASIC PERCENTILES AND CRASH RATES

In terms of safety, crash rates are generally considered one of the best measures for identifying high risk carriers. Figure 4 shows scatter plots of crash rates by percentiles for six BASICs. These crash rates are calculated from the same 473,847 nonparticipating carriers with recent activity in 44 States and the District of Columbia used in Table 24. That is, they are 18-month crash rates calculated from February 2008 through July 2009 using the MCMIS Crash file dated June, 2010. Some of the crash rates are presented on the natural log scale, and similarly, some of the BASIC percentiles are presented on the log scale. The method that produced the strongest association was used.

Each plot is based on 90 data points representing percentiles 11–100. The first 10 percentiles were omitted from the graphs since BASIC percentiles are often not recorded at those small levels, resulting in unstable rates. As an example, for the Unsafe Driving BASIC, the crash rate at the 80th percentile is calculated by aggregating crashes and power units for all carriers that have a BASIC percentile in the interval (79, 80). Using this method of aggregation, each data point represents a group crash rate calculated from generally several hundred carriers.

All of the BASIC measures have positive associations with crash rates, except for two. The Unsafe Driving BASIC has a strong and consistent linear association with crash rates. Of all the BASICs, it is the strongest. Although there is a little more scatter about the regression line, the Fatigued Driving BASIC is also strongly correlated with crash rates. The Vehicle Maintenance BASIC also shows a positive association with crash rates. Note from Table 24 that Vehicle Maintenance and Fatigued Driving BASIC thresholds are the most likely to be exceeded. Although somewhat weaker than the others, the Controlled Substance and Alcohol BASIC has a

positive association with crash rates, but relatively few carriers exceed this BASIC threshold. The greater scatter about the regression line for this BASIC is likely due to smaller sample size.

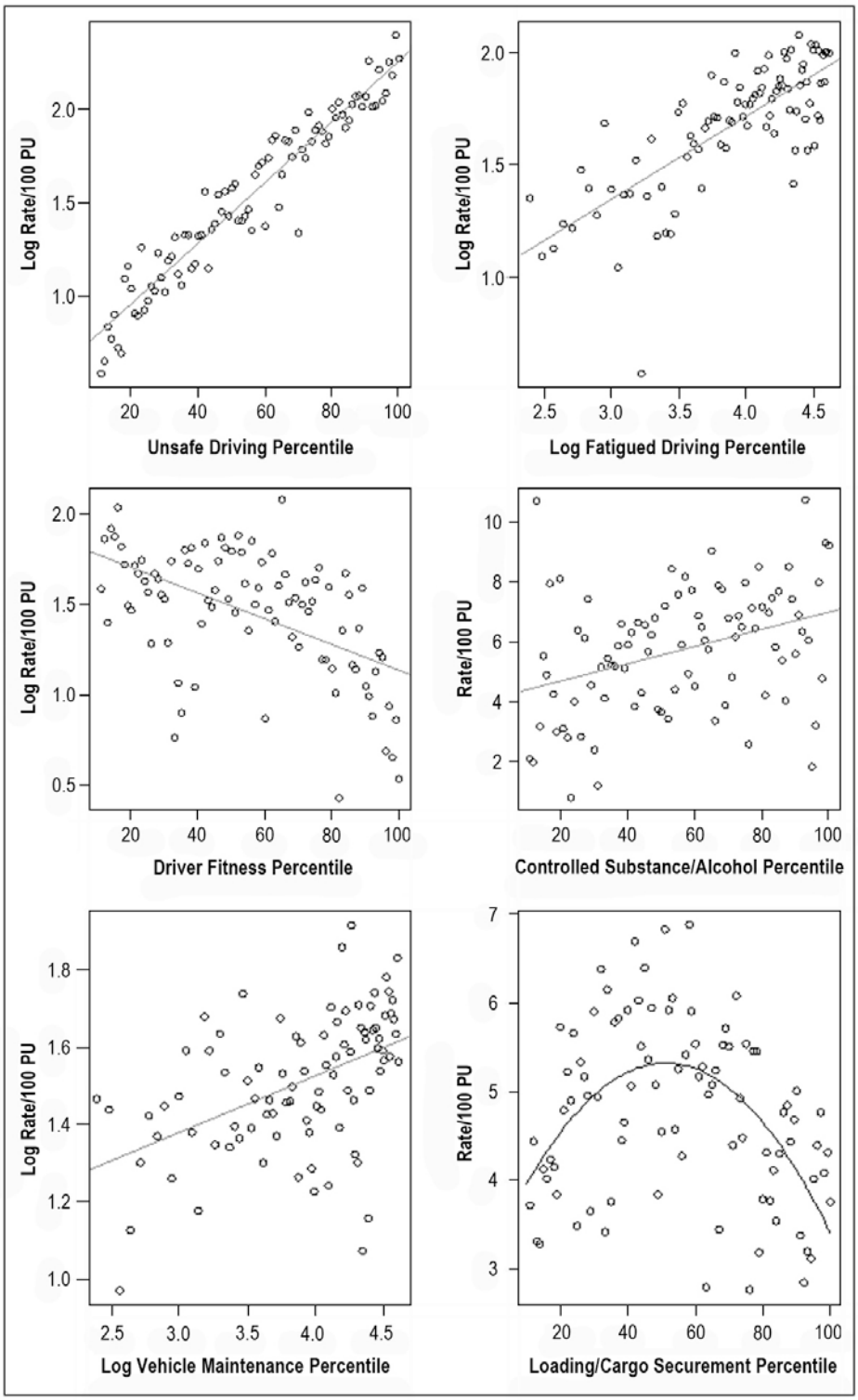

Figure 4. Scatter Plots of Crash Rates by BASIC Percentiles (February 2008–July 2009, MCMIS File Dated June 2010)

The Driver Fitness BASIC shows a negative correlation with crash rates. It may be that this BASIC measures attributes other than those related to crash rates. Table 24 shows that carriers exceeding this BASIC threshold have a crash rate of 3.04 per 100 power units, which is the lowest rate among carriers exceeding any of the other BASIC thresholds. Nevertheless, the ratio to not identified carriers is about 1.5, suggesting that carriers that exceed the Driver Fitness BASIC threshold have a higher crash rate than carriers that do not exceed any BASIC thresholds. These two findings are not in contradiction. One measure compares carriers exceeding a BASIC threshold to carriers that do not exceed any BASIC thresholds. The other is a scatter plot of crash rates ranked by percentile scores. Table 29, however, suggests that there is no difference in crash rates between carriers that exceed the Driver Fitness BASIC and those that do not.

Finally, Figure 5 shows a scatter plot of crash rates by the Crash Indicator BASIC percentiles. As expected, this indicator has the strongest association with crash rates among all measures since it is calculated from crash history. The association is very strong, but note the greater scatter about the regression line at the lower percentiles, say, less than 40. This is due to fewer carriers having crash rates calculated at these percentile values. In other words, these data points are based on fewer carriers, resulting in greater variability.

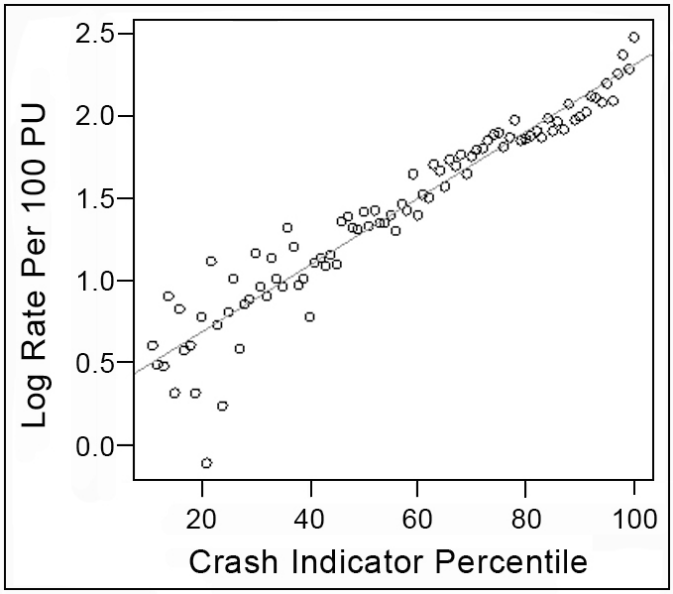

Figure 5. Scatter Plot and Fitted Line of Log Crash Rate by Crash Indicator BASIC Percentile

Figure 6 is a matrix scatter plot of crash rates for each BASIC for those crash rates presented on the log scale. Each plot is based on the 90 data points corresponding to percentiles 11–100. Both the horizontal and vertical axes are log rates. In a matrix scatter plot, the plots in the upper diagonal are the same as those in the lower diagonal with the horizontal and vertical axes reversed. The plots show a strong correlation between the Unsafe Driving BASIC and the Crash Indicator BASIC. The Fatigued Driving BASIC also has log rates that are correlated with Unsafe Driving and Crash Indicator rates. The Vehicle Maintenance rates are associated with the Crash Indicator rates, and others. The Driver Fitness rates display a negative correlation with rates from the other measures.

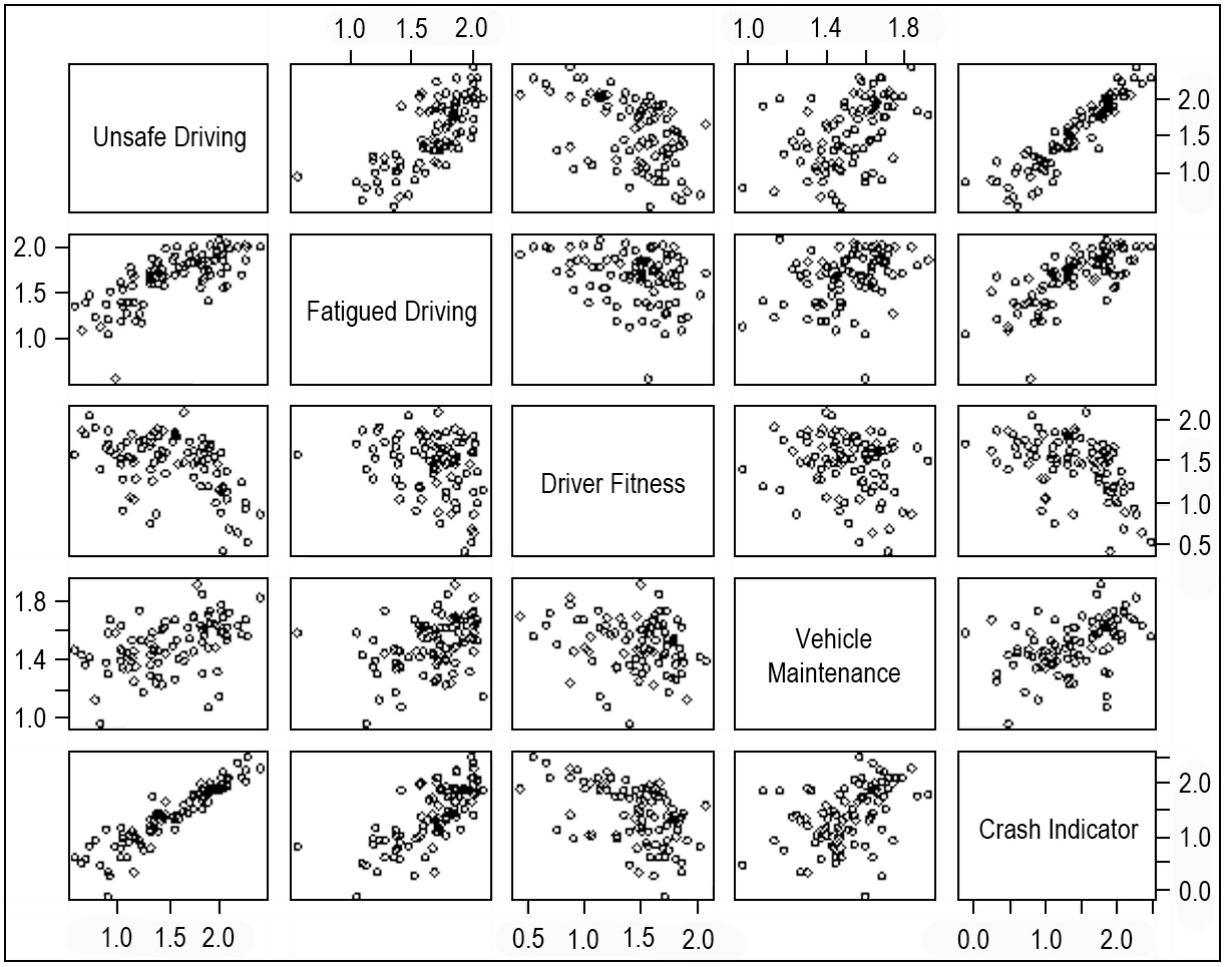

Figure 6. Matrix Scatter Plot of Log Crash Rates for Five BASICs

5.8 A STATISTICAL MODEL FOR PREDICTING CRASH RATES FROM BASIC PERCENTILES

While examining scatter plots of crash rates by BASIC percentiles one at a time is a useful exercise, a statistical model can be developed that uses the input from six BASIC measures simultaneously to measure the association with crash rates. An empirical Bayes model that predicts crash rates was developed to provide a measure of safety of a carrier's on-road crash performance. This model will be used, for example, in section 8 when evaluating the effectiveness of certain interventions by matching a control group's distribution of crash rates to a test group's distribution of crash rates, with the hope that the only difference between the test group and the control group before comparison and during the followup period is that the test group received CSA 2010 intervention(s) and the control group did not. The model, as presented here, will be described in general terms. For a detailed description of the model and assumptions, the interested reader is referred to Appendix E.

A statistical model using the BASICs to predict crash rates has two purposes. The first is to measure the association of each BASIC with the crash rate, while holding all the other BASICs

constant. This provides the best estimate of the association of each BASIC with the fundamental metric of safety, which is the crash rate. The second reason, or rather use, of the model is to be able to provide a stable metric of safety for each carrier, even when a crash rate for a carrier might not be available. In comparison with inspections and the other data the BASICs are based on, crashes are rare events. In fact, the very design of the CSA 2010 is an effort to get a more accurate measurement of carrier safety by incorporating all available data. Since crashes are relatively rare events, more frequent events such as the inspections and violations that go into the BASICs are used to predict crash rates. Because they are based on more common and frequently measured events related to safety, a statistical model using the BASICs can be a more stable predictor of carrier safety than the crash rate itself.

The model predicts crash rates using six BASIC percentile measures and the carrier's operation type (passenger, hazardous material, general). The Crash Indicator BASIC is not included as a predictor variable in the model since the crash rate is being modeled as the response variable. As throughout this section, and to be consistent with analyses already presented, input data were taken from the 473,847 nonparticipating carriers. Crash rates were calculated over the 18-month period beginning with the start of the Operational Model Test in February 2008 and ending in July 2009. The MCMIS Crash data that are used as input to the SMS are dated June 25, 2010. Therefore, the crash data used in the empirical Bayes model should be representative of data that eventually are uploaded to the MCMIS file since June 2010 is 11 months after the date of any crashes used in the statistical model.

One major difference between this analysis and previous analyses of crash rates is that this analysis is performed at the carrier level. Previous analyses used aggregated data to calculate crash rates for *groups* of carriers. For this analysis, each record corresponds to one carrier, and the model predicts a crash rate based on the BASIC percentiles for a particular carrier.

The 473,847 carriers provide a large sample to validate the empirical Bayes model. Because the sample is so large, a procedure was used to select a stratified random sample of 9,500 carriers using the number of power units and the carrier's operation type as strata. Table 34 shows the 437,847 carriers categorized by power units and operation type. The lower portion of the table shows the sample of 9,500 categorized by the 15 strata. Where sample sizes are relatively small, all carriers were selected. For example, all 574 carriers with more than 500 power units were included. In addition, all passenger carriers with more than 50 power units were included. The resulting sample includes 1,500 passenger carriers, 3,000 hazardous material carriers, and 5,000 general carriers. The idea is to select a representative sample of carriers and to design a model that can predict a crash rate for a carrier regardless of the number of power units and the operation type.

If a carrier does not meet the data sufficiency requirements of a particular BASIC, a percentile score is not calculated. Few carriers have percentile scores recorded for all BASICs, let alone for even a few. For this reason, a categorical variable was created for inclusion in the model. For each of the six BASICs, the variables were categorized into three levels:

- 1 = percentile not recorded.
- 2 = percentile recorded, but below BASIC threshold.

- 3 = percentile recorded and above BASIC threshold.

In this way, carriers that have percentiles recorded for any of the BASICs, or none at all, can be included in the model. When predictor variables are categorical, one level, usually the first, is designated as the *baseline* category, and the other levels are fit as predictor variables. In this model, the first level, where the percentile is not recorded, will serve as the baseline or referent category. Since the three levels have a natural ordering, the usual convention is to set the first level as the baseline case. The other two levels of the variable, levels 2 and 3, will be fit in the model as predictor variables.

Table 34. Population and Sample Size of Carriers Used in the Empirical Bayes Model

Power Units	Passenger	Hazmat	General	Total
1–-5	2,275	18,463	376,510	397,248
6–15	736	7,338	43,393	51,467
16–50	413	4,135	13,698	18,246
51–500	206	2,329	3,777	6,312
>500	12	295	267	574
Total	**3,642**	**32,560**	**437,645**	**473,847**

Table 35. Sample Size of Carriers Used in the Empirical Bayes Model

Power Units	Passenger	Hazmat	General	Total
1–5	500	900	1,500	2,900
6–15	400	700	1,233	2,333
16–50	382	605	1,000	1,987
51–500	206	500	1,000	1,706
>500	12	295	267	574
Total	**1,500**	**3,000**	**5,000**	**9,500**

Figure 7 shows the empirical Bayes estimator of the crash rate for carrier *i*.

$$\frac{y_i + 1/\alpha}{t_i + 1/(\alpha\mu_i)} = \left(\frac{t_i \alpha\mu_i}{1 + t_i \alpha\mu_i}\right)\frac{y_i}{t_i} + \left(\frac{1}{1 + t_i \alpha\mu_i}\right)\mu_i$$

where
y_i = number of crashes
t_i = number of power units
μ_i = model-based rate calculated from negative binomial regression
a = negative binomial scale parameter

Figure 7. Equation for the Empirical Bayes Estimator of the Crash Rate for Carrier *i*

For a particular carrier, *y/t* is the observed crash rate and *mu* is the rate predicted by a negative binomial regression model. Therefore, the Bayes estimator is a weighted average of the observed

rate and the rate predicted by negative binomial regression. When a carrier has a lot of power units, the observed rate tends to be a good estimator and the Bayes estimator *smoothes* towards the observed rate. When a carrier has few power units, the observed rate is not a good estimator and the Bayes estimator smoothes towards the negative binomial rate, which borrows information from all carriers through a regression model on BASIC percentiles. The scale parameter *alpha* also helps to determine how much weight the observed rate receives, and how much weight the negative binomial rate receives. As *alpha* tends to zero, more weight is given to the negative binomial rate.

In addition to the six BASIC measures, the carrier's operation type is also included in the regression model. It is a categorical variable with three levels:

- 1 = passenger carrier.
- 2 = hazardous material carrier.
- 3 = general carrier.

The first level, passenger carrier, will serve as the baseline category and the other two levels, hazardous material and general, will be fit as variables in the negative binomial regression.

The regression model is fit as a log-linear model. If *beta* represents regression parameters and *chi* represents predictor variables, the negative binomial rates *mu* are calculated by exponentiation of the regression function as shown in Figure 8.

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \cdots\cdots + \beta_p x_{ip})$$

Figure 8. Equation for Calculating the Negative Binomial Rates *Mu*

Table 36. Fit of a Log-linear Negative Binomial Regression Model to Crash Data shows the fit of a negative binomial regression on six BASIC measures and operation type. Note that the baseline level is set to zero and is not fit. Since this is a log-linear model, the parameters have interpretations as log relative risks, relative to the baseline category. Relative risks can be estimated by exponentiation of parameter estimates. Positive coefficients are associated with increased risk, while negative coefficients are associated with reduced risk. For example, the crash rate for a carrier that exceeds the Unsafe Driving threshold is approximately exp(1.0284) = 2.8 times the rate of a carrier that has no percentile measured for this BASIC.

The magnitudes of the chi-square statistic can be used to assess significance in the model. In addition, p-values less than 0.05 are significant at the usual 0.05 level. After Unsafe Driving, Vehicle Maintenance has the largest chi-square statistic. The crash rate for a carrier exceeding the Vehicle Maintenance threshold is approximately exp(0.5816) = 1.8 times the rate of a carrier that has no percentile measured for this BASIC. Other estimates can be judged similarly. It appears that Unsafe Driving, Vehicle Maintenance, and Fatigued Driving are the three most important predictors in the model.

The signs of the parameters in the model tend to be consistent with the scatter plots shown in Figure 6. Note the negative coefficients attached to Driver Fitness, which support the scatter plot showing a negative correlation with crash rates. In addition, the Substance/Alcohol variable is not significant in this model since the p-values are greater than 0.05. At the third level, the Loading/Cargo variable is technically not significant, with a p-value of 0.0764. The negative coefficients attached to operation type suggest that passenger carriers have a higher crash rate than either hazardous material or general carriers. Hazardous material carriers have the lowest crash rate among operation types.

Table 36. Fit of a Log-Linear Negative Binomial Regression Model to Crash Data

Parameter	Level	Estimate	Std Err	Chi-square	p-value
Intercept	–	-4.0006	0.0564	5,026.4	<.0001
Operation Type	2	-0.4628	0.0661	49.0	<.0001
Operation Type	3	-0.4247	0.0610	48.4	<.0001
Unsafe Driving	2	0.4259	0.0581	53.7	<.0001
Unsafe Driving	3	1.0284	0.0755	185.6	<.0001
Fatigued Driving	2	0.3063	0.0563	29.6	<.0001
Fatigued driving	3	0.3499	0.0789	19.7	<.0001
Driver Fitness	2	-0.1903	0.0581	10.7	0.0011
Driver Fitness	3	-0.3538	0.0726	23.8	<.0001
Substance/Alcohol	2	-0.0620	0.0893	0.5	0.4877
Substance/Alcohol	3	-0.0312	0.1086	0.1	0.7741
Vehicle Maintenance	2	0.4781	0.0576	68.9	<.0001
Vehicle maintenance	3	0.5816	0.0764	57.9	<.0001
Loading/Cargo	2	0.2107	0.0605	12.1	0.0005
Loading/Cargo	3	0.1174	0.0662	3.1	0.0764
Scale α	–	0.9663	0.0348	–	–

It is important to note that the negative binomial regression results are only part of the Bayes estimator used here. The Bayes estimator uses both the observed rate and the predicted rate under the regression model since it compromises somewhere in between. It uses the good properties of each estimator to determine which one receives more weight. For a particular carrier, whichever estimate is more appropriate gets used.

As an example of the properties of the Bayes estimator, Table 37 shows a comparison of the estimated empirical Bayes rates for two carriers. Carrier 1 has only one crash and one power unit. Therefore, the observed rate is 100 crashes per 100 power units. Due to the small number of crashes and power units, this rate suffers from uncertainty due to large variation. The negative binomial rate is based on a regression model that combines the information from all carriers that have a similar number of power units and similar BASIC thresholds exceeded. Note that this carrier exceeded the Unsafe Driving threshold, which is highly correlated with crash occurrence. The predicted rate from the negative binomial regression is 5.40. The Bayes rate is 10.09 and compromises between the observed rate and the negative binomial rate, but in this case gives more weight to the negative binomial rate. On the other hand, Carrier 2 has 14 crashes and 134

power units, so the observed rate is a good estimator. In this case, the Bayes estimator will give almost all weight to the observed rate, since the observed is based on relatively many power units, and is therefore stable.

Table 37. Comparison of Empirical Bayes Rates for Two Carriers

Carrier	BASIC Thresholds Exceeded	Crashes (y)	Power Units (t)	Observed Rate (y/t) x100	Negative Binomial Rate μ	Empirical Bayes Rate
Carrier 1	Unsafe Driving	1	1	100.00	5.40	10.09
Carrier 2	None	14	134	10.45	4.02	9.39

6. INTERVENTION CYCLES

This section identifies the cycles of interventions that were applied to the test carriers in the four primary test States, establishes the primary patterns of interventions that were applied, and evaluates the association of the primary patterns with the underlying safety of the carriers. The underlying safety of the carriers is characterized in two ways: in terms of the number of BASICs exceeded by the carriers and in terms of the crash rate of the carrier at the initiation of the intervention cycle. The section also begins the process of evaluating the effectiveness of interventions by analyzing the extent to which it was necessary to follow up one intervention type with another.

6.1 IDENTIFICATION OF INTERVENTION CYCLES

Intervention "cycles" can be identified in the data, although they are not explicitly coded as such. Many carriers are subject to more than one intervention. If the initial intervention does not result in the intended improvement, additional intervention(s) are imposed. In the data, there is an indication of which interventions are part of these cycles. The intervention data table includes a field ("Inter_link") that lists interventions that are "linked" together.

Using the Inter_link field, sets of linked interventions, which will be referred to as cycles of interventions, were established. Within each cycle, the initial BASIC percentiles are recorded along with the date of the first intervention, the type of each intervention in the cycle, and the date when the final intervention in the series was marked closed/completed (all tasks completed and the intervention marked as closed), if the last intervention was closed.

It was noted during the process of constructing the intervention cycles that a warning letter intervention is never linked explicitly to a subsequent intervention; i.e., when the intervention is a warning letter, the Inter_link field is always left blank. Yet, review of the series of interventions received by a carrier makes it clear that warning letters are often followed up by other intervention types. Accordingly, warning letters were added to intervention cycles when the warning letter was followed within 12 months by another intervention. Twelve months was set as the interval because the CSA training manual suggests that the investigators monitor the carriers at least 6 months following a warning letter to allow them time to improve. If a carrier's safety does not improve within 12 months after a warning letter, then follow up is necessary. Accordingly, if another intervention followed a warning letter within 12 months, the warning letter was included in the intervention cycle for the carrier.

Intervention cycles permit a more detailed look at the effectiveness of the interventions in the context in which they are actually applied in the field. As carriers are identified and intervened upon, some will improve and the action is closed, while others may not improve and further intervention is necessary to improve the carrier's compliance and safety outcome. An intervention type that was always followed by another intervention could be considered ineffective, since the intervention only rarely resolved the issue.

Some types of interventions can be considered and often are "terminal," such as the NOV or NOC, in that they culminate a series of interventions aimed at altering behavior by moving

toward a penalty. A CSP is also typically the end point of an intervention cycle, as it involves a carrier developing and implementing a plan to eliminate the high BASIC scores that prompted it. Other types of interventions are clearly part of a graded system of escalation intended to move carriers to compliance and safer behavior. These include the warning letter, offsite investigation, onsite focused investigation, and onsite comprehensive investigation. However, it should be noted that there is no fixed sequence of interventions that must be followed. The CSA 2010 program has a built-in flexibility that permits the intervention selected to be tailored to the circumstances.

For the current analysis, intervention cycles initiated in the first full year of Phase II in the four test States are included. In addition, only cycles in which the last intervention in the series is marked closed/completed at some point before the end of the data collection period in June, 2010, are included. The motivation for these two restrictions is to provide a realistic examination of the intervention process as it will be conducted when the full CSA 2010 program is implemented. During the first phase of the Operational Model Test, A/B carriers were excluded and not all intervention types were available. In Phase II, all test carriers are included and all intervention types are available to the safety investigator)

The goal of the restriction to intervention cycles that were initiated in the first full year of Phase II is to identify a set of intervention cycles that have effectively run their course. At any given point in time, it is not possible to know whether the cycle has come to an end, meaning that the safety and compliance issues have been resolved. The status of each intervention in the cycle is known, including whether or not the intervention has a "closed/completed" status. But the linked set of interventions applied to the carrier as such is not marked as "completed" or "not completed". And in a sense, the books are not closed on an intervention cycle at the point when the last in the list is marked closed/completed because the last stage in the process is to monitor carriers after an intervention is closed for 12 months. The cycle is really an artificial construct, but it is motivated by the fact that the intervention process may consist of several discreet actions, which may be escalated to a different intervention if the required improvement on the part of the carrier does not occur. It is important to identify and evaluate these cycles because they are fundamental to the CSA process.

Table 38 shows the number of intervention cycles experienced by carriers in the first year of Phase II of the Operational Model. Intervention cycles were initiated on a total of 1,964 carriers in the first full year of Phase II, which began in October 2008, and the last intervention in the cycle was marked closed/completed by the end of data collection of the operational model in June 2010. Ultimately, 1,926 or 98.1 percent of those carriers experienced only one intervention cycle, while 36 went through two cycles and two went through three cycles. Accordingly, it is important to keep in mind that, while some few carriers had more than one cycle, the overwhelming majority of carriers analyzed went through, from start to finish, one intervention cycle.

Table 38. Number of Intervention Cycles Per Carrier First Year of Operational Model Phase II

Number of Cycles	Carriers	Percent
1	1,926	98.1%
2	36	1.8%
3	2	0.1%
Total	**1,964**	**100.0%**

Table 39 shows the distribution of the specific intervention types included in the cycles initiated in the first year of Phase II of the Operational Model Test. The warning letter is the most common intervention type, accounting for about 35 percent of the total interventions. An onsite focused was the second most common with about 20.4 percent of the interventions. offsite and onsite comprehensive investigations each accounted for about 11 to 12 percent of the interventions, as did the CSP. NOVs were fairly rare, with only 1.1 percent. An NOV is a formal notice of safety deficiencies that requires a response from the carrier. The violations are severe enough to warrant formal action, but not a civil penalty such as a fine. The NOC is an issuance of a civil fine for severe violations of the FMCSR. Almost 9 percent of the intervention cycles initiated in the first year of Phase II culminated in an NOC.

Table 39. Interventions in Cycles Initiated in First Year of Operational Model Phase II

Intervention Type	N	Percent
Warning Letters	1,104	34.6%
CSP	382	12.0%
NOV	35	1.1%
NOC	278	8.7%
Offsite	375	11.7%
Onsite Focused	650	20.4%
Onsite Comprehensive	369	11.6%
Total	**3,193**	**100.0%**

More than 50 percent of intervention cycles consisted of only one intervention, and another 37.1 percent consisted of only two interventions. Overwhelmingly, the intervention cycles consist of only one or two interventions. Table 40 also shows the relationship between the number of interventions in a cycle and the number of BASICs exceeded. The association is in the expected direction. Also shown is the predicted crash rate using results from the statistical model presented in Section 8, and again the relationship is in the expected direction. At the initiation of the intervention cycle, carriers that require more interventions to improve, tend to have exceeded more BASICs and to have higher predicted crash rates. Note also that 99.0 percent of the cycles are accounted for by three or fewer interventions. More than half include only one intervention, about 37 percent had only two interventions and an additional 9.5 percent included three interventions. More than three interventions in a cycle is quite rare, with only about 1 percent of the carriers who were intervened upon receiving more than three interventions. These carriers also exceeded more BASIC thresholds and had significantly higher crash rates, as predicted by the statistical model. In fact, the predicted crash rates for carriers with intervention cycles that include four or more interventions are more than twice as high as those who took fewer interventions to improve.

Table 40. Number of Interventions in the Intervention Cycles, Mean Number of BASICs Failed and Initial Crash Rate

Number of Interventions	Number of Intervention Cycles	Percent	Mean # BASICs Failed	Mean Crash Rate Predicted	Median Crash Rate Predicted
1	1,042	52.4	1.22	3.6	2.5
2	738	37.1	1.45	4.6	2.9
3	188	9.5	1.28	4.0	2.8
4	14	0.7	2.00	9.7	3.4
5	2	0.1	2.00	6.0	6.0
6	1	0.1	2.00	11.9	11.9
7	3	0.2	2.67	6.9	4.8
Total	**1,988**	**100.0**	**1.32**	**4.0**	**2.7**

There is also a relationship between the nature of the first intervention in a cycle and the relative safety of the carrier as measured by the predicted crash rate (Figure 9). Intervention cycles that were initiated by a warning letter tended to be for carriers that had lower crash rates than those where the initial intervention in the cycle was one of the more severe, such as an investigation, whether offsite, onsite focused, or onsite comprehensive. Carriers in which the first intervention was a warning letter had an average predicted crash rate of 3.1, compared with 3.9 for those in which the intervention cycle started with an offsite, 5.0 where it was an onsite focused, and a mean predicted crash rate of 6.2 where the intervention cycle started off with an onsite comprehensive investigation. The figure shows the measure of safety (predicted crash rate) by the type of initial intervention in the intervention cycle for the four most common initial intervention types. (Two cycles were initiated by a NOV and two by a NOC. These cycles are not shown.) The figure also shows 95 percent confidence intervals for the predicted crash rates.

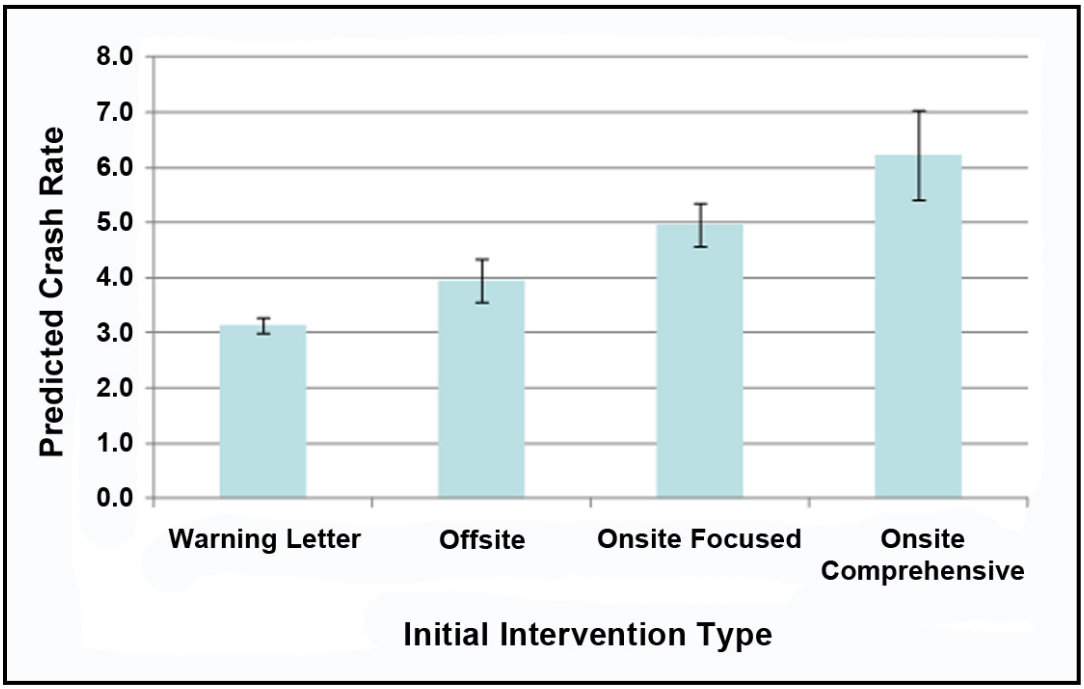

Figure 9. Predicted Crash Rate by First Intervention Type

This pattern is expected since interventions are escalated for carriers with more serious or more frequent violations, and carriers that exceed multiple BASICs tend to have higher crash rates. For example, carriers that have exceeded two or more BASICs, including the Fatigued Driving or Crash Indicator BASIC, are recommended by the CSA process to receive an onsite investigation. Carriers that have not exceeded the Fatigued Driving or Crash Indicator BASIC but have had a prior failed BASIC are recommended for an offsite investigation. Carriers that are repeat offenders or who have exceeded the critical Fatigued Driving or Crash Indicator BASIC have lower levels of safety, as measured by the predicted crash rate. Thus, the rules guiding the selection of appropriate interventions seem to be correctly targeting carriers with high crash rates.

6.2 INTERVENTION PATTERNS

One measure of the effectiveness of individual intervention types is whether the intervention resolved the issue or whether further interventions were necessary. An intervention that was followed by safety improvement and no further intervention required would be considered effective, but a type where it was almost always necessary to escalate to another intervention would not be considered effective. This is just one measure of effectiveness. Section 5 will consider each intervention type and measure the change in safety following interventions, both for individual interventions and patterns of interventions.

Table 41 shows the number of interventions that followed each intervention type in the intervention cycles. Within each intervention cycle, the first instance of each type was identified and then the subsequent interventions were counted. The top number shows the frequency counts and the bottom number shows the percentages for each intervention type.

It appears that each of the intervention types is effective in some situations. In 61.7 percent of the intervention cycles in which a warning letter was issued, there were no followup interventions. In almost all of these cases, of course, the warning letter was the first intervention in the cycle. And in most cases, it appears that the warning letter was sufficient. In about 24 percent of the cases, one further intervention was necessary, and in an additional 14 percent, two more interventions were needed. Each of the other primary intervention types also was effective in 45 to 55 percent of the cycles, and additional interventions were needed in the remainder. Note that most intervention cycles consisted of only one or two interventions and that carriers requiring more interventions tended to exceed more BASIC thresholds and to have higher crash rates (Table 40).

Table 41. Number of Subsequent Interventions, by Intervention Type

Intervention Type	0	1	2	3	4	5	6	Total
Warning Letter	680 61.7%	261 23.7%	153 13.9%	7 0.6%	1 0.1%	0 0.0%	1 0.1%	1,103 100%
Offsite	166 44.6%	196 52.7%	9 2.4%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	372 100%
Onsite focused	349 54.5%	263 41.1%	23 3.6%	3 0.5%	0 0.0%	2 0.3%	0 0.0%	640 100%
Onsite Comprehensive	164 45.6%	179 49.7%	10 2.8%	4 1.1%	1 0.3%	0 0.0%	2 0.6%	360 100%
CSP	367 96.1%	13 3.4%	1 0.3%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	382 100%
NOV	29 85.3%	5 14.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	34 100%
NOC	202 87.1%	18 7.8%	8 3.4%	2 0.9%	2 0.9%	0 0.0%	0 0.0%	232 100%

Certain intervention types—CSP, NOV, NOC—in most cases are not followed by further interventions. In more than 96 percent of instances in which a CSP is implemented there is no follow-on intervention. This is evidence that the CSP is an effective intervention. In most of the NOVs, the following intervention is a NOC, and in the case of the NOCs, where there is more than one, the following is almost always another NOC.

Table 42 shows the most frequent patterns of interventions in the first full year of Phase II of the Operational Model Test. The patterns are sorted from the most frequent to the least frequent, and only the top 16 combinations are shown (there were 79 different patterns) for reasons of space. These top 16 account for 91 percent of the intervention cycles. Note that in these 16 patterns, none included four or more interventions. As is noted below, most intervention cycles include only one or two types. More than three interventions in a cycle is rare. The N column presents counts of intervention cycles for carriers. As noted, in almost all cases carriers experienced only one cycle of interventions. One-third of the cycles were completed by issuing a warning letter only. Another 9 percent consisted of only an onsite focused investigation, and about 7 percent included only a warning letter, followed by an onsite focused investigation.

Table 42. Primary Intervention Patterns, Cycles Initiated in First Year of Phase II

1st Intervention	2nd Intervention	3rd	4th	5th	N	Percent
Warning Letter	None	None	None	None	688	33.3%
Onsite Focused	None	None	None	None	180	9.0%
Warning Letter	Onsite Focused	None	None	None	145	7.2%
Onsite comprehensive	None	None	None	None	130	6.5%
Onsite Focused	CSP	None	None	None	125	6.2%
Offsite	CSP	None	None	None	92	4.6%
Warning Letter	Offsite	None	None	None	88	4.4%
Onsite Comprehensive	NOC	None	None	None	80	4.0%
Offsite	None	None	None	None	72	3.6%
Onsite Comprehensive	CSP	None	None	None	49	2.4%
Warning Letter	Offsite	CSP	None	None	47	2.3%
Onsite Focused	NOC	None	None	None	43	2.1%
Warning Letter	Onsite Focused	CSP	None	None	40	2.0%
Warning Letter	Onsite Focused	NOC	None	None	26	1.3%
Warning Letter	Onsite Comprehensive	None	None	None	19	0.9%
Warning Letter	Onsite Comprehensive	NOC	None	None	16	0.8%

The intervention patterns in the cycles are clearly associated with the underlying safety of the carriers. This result is consistent with the results above, but Table 42 illustrates the point. The table shows the mean crash rate for each of the 16 intervention cycles identified in Table 42. (The N column shows the number of cycles for which a valid crash rate could be calculated, not the total number of such cycles, as in Table 42.) Cycles resolved only by a warning letter have the lowest crash rate. The next most frequent pattern, an onsite focused investigation with no follow up, has a substantially higher crash rate at 4.9. The warning letter followed by an onsite focused has a lower crash rate, but still significantly higher than the pattern where the warning letter was sufficient. The highest crash rates occur for patterns that include an onsite comprehensive investigation and some further intervention. As the prior analysis has shown, the number of interventions and the severity of interventions are strongly associated with the crash rate.

Table 43. Mean Crash Rate by Primary Intervention Patterns, Intervention Cycles Initiated in First Year of Phase II

1st	2nd	3rd	N	Mean Crash Rate
Warning Letter	None	None	667	2.9
Onsite Focused	None	None	174	4.9
Warning Letter	Onsite Focused	None	145	3.6
Onsite Comprehensive	None	None	125	4.9
Onsite Focused	CSP	None	125	4.6
Offsite	CSP	None	92	4.2
Warning Letter	Offsite	None	88	3.4
Onsite Comprehensive	NOC	None	79	6.2
Offsite	None	None	72	4.1
Onsite Comprehensive	CSP	None	49	7.5
Warning Letter	Offsite	CSP	47	3.2
Onsite Focused	NOC	None	41	4.4
Warning Letter	Onsite Focused	CSP	40	3.4
Warning Letter	Onsite Focused	NOC	26	2.9
Warning Letter	Onsite Comprehensive	None	19	4.3
Warning Letter	Onsite Comprehensive	NOC	16	4.5

7. EFFECT OF WARNING LETTERS ON BASIC PERCENTILES

The effect of interventions on BASIC percentiles is explored by comparing test carriers and control carriers in the original four States that exceeded a particular BASIC threshold for the first time, and then following those carriers for 12 months after exceeding the threshold. The analysis in this section focuses on test carriers that received a warning letter as the only intervention. The reason for this restriction is that the warning letter is typically closed within a month, which facilitates evaluating the effectiveness of the intervention. Carriers with multiple interventions have longer times to closed status after exceeding a BASIC threshold and are less likely to have much followup time for evaluation. In addition, the sample sizes for carriers with multiple interventions are small and are generally too small to make valid inferences about changes in specific BASIC scores even over 6 months, let alone 12. The effectiveness of other types of interventions is considered in section 8, which examines their impact upon motor compliance in general, rather than on specific BASIC scores.

The purpose here is to demonstrate the general effectiveness of the warning letter intervention on BASIC percentiles by comparing test and control carriers that initially exceeded the same BASIC threshold for the first time over a 12-month followup period. The CSA 2010 data provide sufficient sample sizes for this comparison. The control carriers are those in the original four States that had no CR during the Operational Model Test.

Because the test group received the warning letter as the only intervention, the test group can be regarded as a low-risk group of carriers in terms of safety performance. However, the control group carriers had no CRs during the Operational Model Test, and may also be regarded as a mild group among control carriers in terms of safety performance.

The BASICs considered in this analysis include:

- Unsafe Driving.
- Fatigued Driving.
- Driver Fitness.
- Vehicle Maintenance.
- Improper Loading/Cargo Securement.

The Controlled Substance and Alcohol BASIC was not considered due to insufficient data for evaluation. Among the BASICs, the threshold for this one is least likely to be exceeded by carriers. See, for example, Table 8 that shows that this BASIC is rarely exceeded in relation to the other ones.

Key components of the analysis are listed below:

- Both test and control carriers exceeded the same BASIC threshold for the first time.

- During the Operational Model Test, the test carriers received the warning letter as the only intervention and the status of that intervention is closed.
- The date of the warning letter was after the date that the BASIC threshold was exceeded.
- During the Operational Model Test, the control carriers received no CRs under the agency's current enforcement protocols.
- Both test and control carriers were followed for 12 months after exceeding the BASIC threshold and are compared based on the percentage of carriers still exceeding the threshold.
- Not all carriers contribute 12 months of followup time. If a carrier exceeded a threshold for the first time in month 25 of the 29 month evaluation period, that carrier would contribute data from months 26 through 29, resulting in four months of followup data.
- Supporting data for all plots, including sample sizes, are given in Appendix A.

It is recognized that warning letters are often sent to carriers in batches, and not necessarily immediately after a BASIC threshold is exceeded. Therefore, a few months may lapse between the time a carrier exceeds a threshold and the time it receives the intervention. However, the 12 month followup period includes sufficient time after the warning letter has been received by the carrier for evaluation. The reason for starting the evaluation at the time the carrier first exceeds the threshold, and not at the time of intervention, is because control carriers did not receive the intervention. In this way, both test and control carriers are being evaluated over the same 12 month moving window.

Figure 10 shows the percentage of carriers exceeding the Unsafe Driving threshold over a 12-month period. At time zero, both test and control carriers exceeded this BASIC for the first time. Over the next 12 months the two groups are followed and the percentage still exceeding the threshold is shown. The "time zero" point is generally a different month for each carrier. It represents the first month that a carrier exceeded the threshold. In addition, not all carriers have 12 months of followup data. If a carrier has only 3 months of followup data after exceeding a BASIC, it contributes only the first 3 months of data to the plot. Figure 10 shows that after 12 months, 14.5 percent of the test carriers were still exceeding the threshold, compared to 37.4 percent of the control carriers. For the test group, there were initially 132 carriers for evaluation, and 69 carriers had complete data for the entire 12 month period.

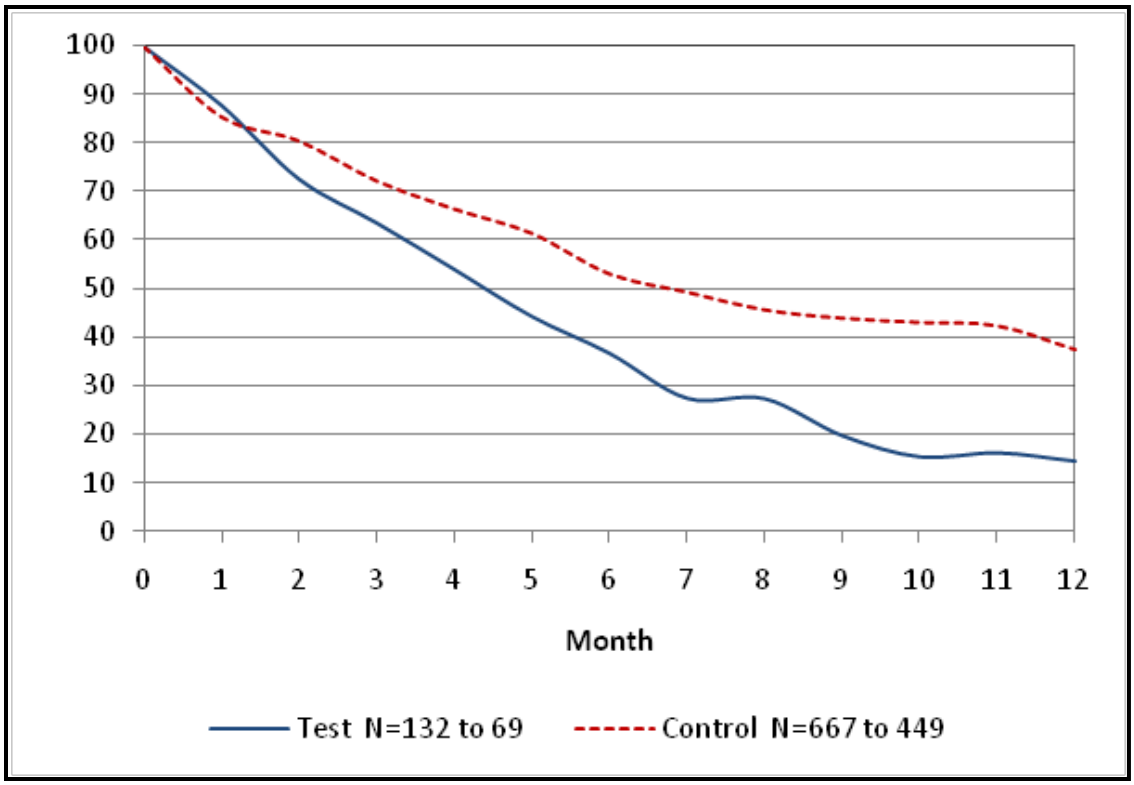

Figure 10. Percent of Carriers Exceeding the Unsafe Driving Threshold With 12 Months of Followup After Exceeding the Threshold (Test Carriers Closed With Warning Letter, Control Carriers With No CR During the Model Test)

The following plots show similar results for the remaining BASICs. Figure 11 and Figure 13 show plots for the Fatigued Driving and the Driver Fitness BASICs, respectively. After the 12-month followup period, 24.6 percent of test carriers and 40.6 percent of control carriers were still exceeding the Fatigued Driving threshold. For the Driver Fitness BASIC, 19.2 percent of test group carriers and 35.8 percent of control group carriers were still exceeding the threshold after 12 months of followup. The number of carriers at the beginning and end of the 12 month period is also shown in each plot.

There appears to be a period of about two months in these plots where there is no difference between test and controls in terms of the percentage of carriers still exceeding the threshold. This could represent the average lag time between a test carrier exceeding a threshold, and the time it receives the warning letter.

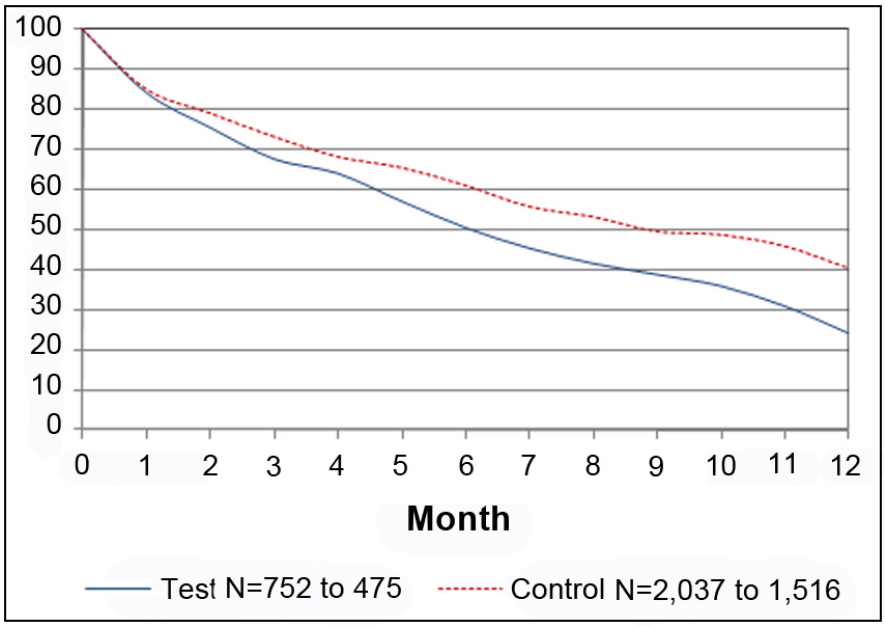

Figure 11. Percent of Carriers Exceeding the Fatigued Driving Threshold With 12 Months of Followup After Exceeding the Threshold (Test Carriers Closed With Warning Letter, Control Carriers With No CR During the Model Test)

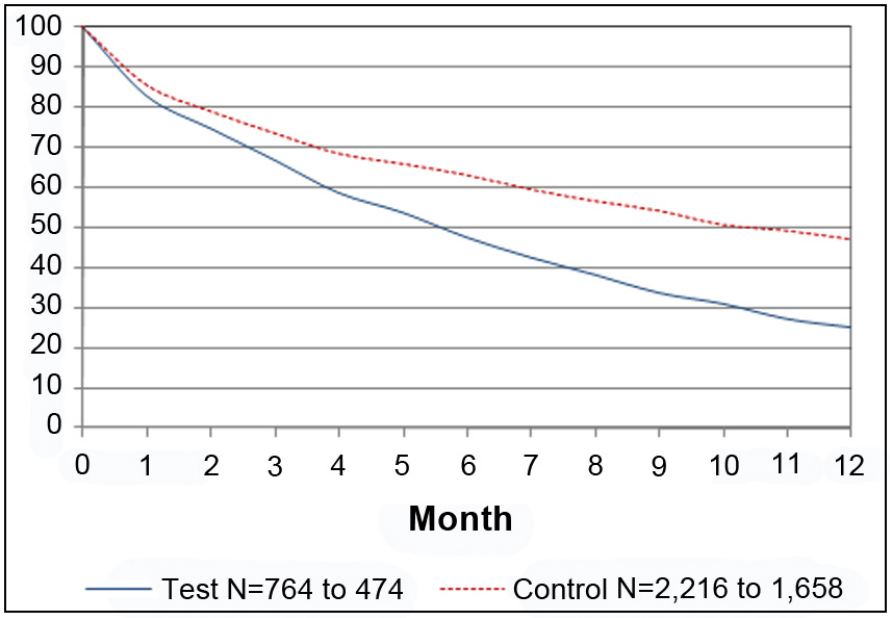

Figure 12. Percent of Carriers Exceeding the Driver Fitness Threshold With 12 Months of Followup After Exceeding the Threshold (Test Carriers Closed With Warning Letter, Control Carriers With No CR During the Model Test)

Figure 13 and Figure 14 show results for the Vehicle Maintenance and the Improper Loading/Cargo Securement BASICs. Again, the test groups show considerable improvement over the control groups. After the 12-month followup period, 25.1 percent of test carriers, and 47.0 percent of control carriers were still exceeding the threshold. For the Improper

Loading/Cargo Securement BASIC, 35.6 percent of test group carriers and 49.9 percent of control group carriers were still exceeding the threshold after 12 months of followup.

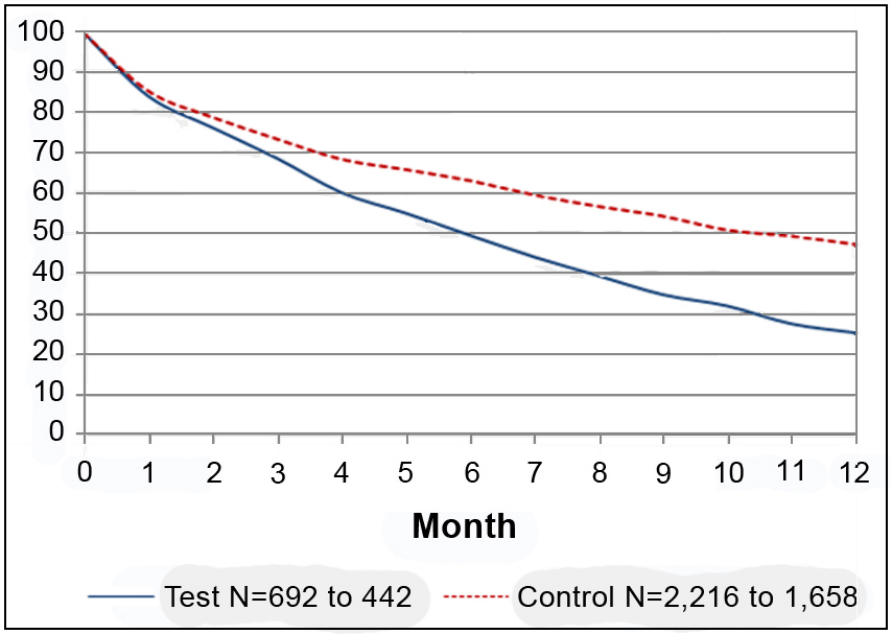

Figure 13. Percent of Carriers Exceeding the Vehicle Maintenance Threshold With 12 Months of Followup After Exceeding the Threshold (Test Carriers Closed With Warning Letter, Control Carriers With No CR During the Model Test)

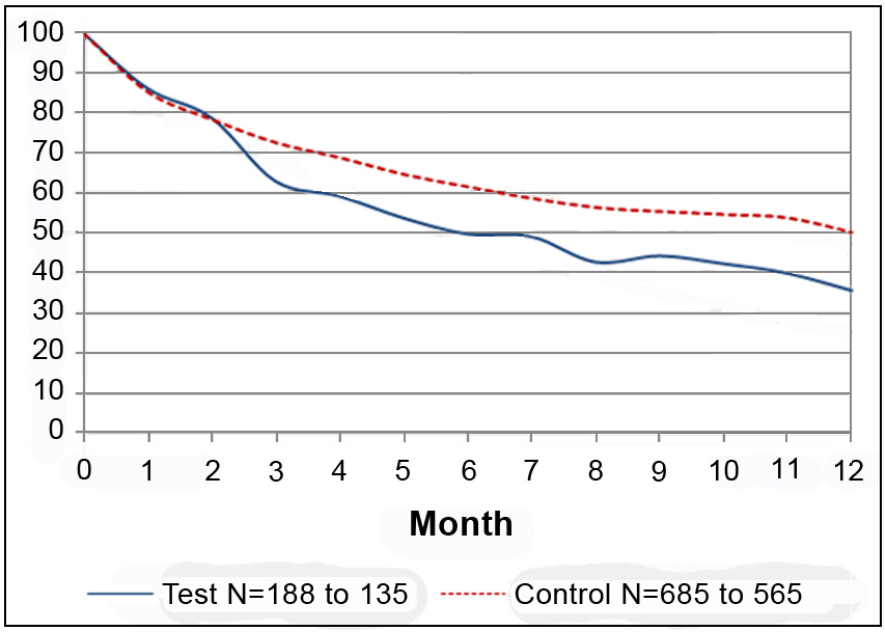

Figure 14. Percent of Carriers Exceeding the Improper Loading/Cargo Securement Threshold With 12 Months of Followup After Exceeding the Threshold (Test Carriers Closed With Warning Letter, Control Carriers With No CR During the Model Test)

[This page intentionally left blank.]

8. EFFECTIVENESS OF CSA 2010 INTERVENTIONS

In this section, the effect of various CSA 2010 interventions on motor carrier compliance is examined by comparing test carriers that received CSA 2010 interventions both to control carriers that did not receive any interventions, and to control carriers that received traditional CRs in the original four States. The test and control group populations are evaluated by examining the percentage of carriers in each group exceeding any of the BASIC thresholds over a 12-month followup period.

Using results from Table 13 and Table 42, intervention patterns were chosen according to those that occurred most frequently and provide sample sizes adequate for analysis. The warning letter followed by no further interventions accounts for about one-third of the intervention cycles and is by far the most common intervention pattern. In total, seven intervention patterns are considered and evaluations are presented in the following order:

- One Intervention.
  - Warning Letter.
  - Offsite Investigation.
  - Onsite Focused Investigation.
  - Onsite Comprehensive Investigation.
- Multiple Interventions.
  - Warning Letter—Any Investigation.
  - Warning Letter—Any Investigation—CSP.
  - Any Intervention Process Terminated by an NOC.

The choice of a good control group is necessary in order to make the comparison valid. For example, there are several thousand control carriers that can be selected for comparison to the test carriers in the four original States. However, the control group should be as similar as possible to the test group being evaluated in terms of several key factors so that the only difference between the two groups before evaluation begins is that the test group received CSA 2010 intervention(s), while the control group did not. Then, any difference in safety performance between the two groups in a followup period is likely attributable to the CSA 2010 intervention(s).

As an example, suppose it is of interest to compare test carriers that exceeded at least one SMS threshold, received a warning letter as the first intervention, were escalated to one of the investigations (offsite, onsite focused, onsite comprehensive) for a second intervention, and finally agreed to a CSP for yet a third intervention to a group of control carriers. Based on the CSA 2010 methodology, the warning letter was not sufficient enough to bring the test carrier's SMS percentiles below acceptable thresholds since the intervention process was escalated to an investigation resulting in a CSP. Therefore, this particular group of test carriers has likely exhibited safety behavior that is considered to be at a higher level of risk than the general population of control carriers. A comparison of this test group to a general population of control carriers would not account for differences in levels of safety at the beginning of the evaluation,

and it would be difficult to determine if any differences in safety performance in a followup period were due to CSA 2010 interventions received by the test group.

To account for differences in levels of safety between test and control groups, several key factors were matched between the two groups prior to evaluation. Figure 15 is a timeline describing the procedure used for comparison. First, both test and control carriers must exceed at least one SMS threshold for the first time. This triggers the intervention process for the test group and puts both test and control carriers in the group of approximately 6.1 percent of carriers (9.9 percent of carriers with recent activity) that exceed at least one threshold. The control group is selected such that its crash rate distribution closely matches the distribution of the crash rate for the test group for the period 2 years prior to exceeding the first threshold. Crash rates are predicted using a statistical model described at a high level in section 5.8. Since control carriers do not receive CSA 2010 interventions, they were randomly assigned a time to closed/completed (recall that closed/completed indicates that all tasks associated with the specific interventions have been accomplished and the intervention moved to a closed status) that matches the distribution of time to closed/completed for the test carriers. This was done by sampling from a normal distribution with the mean and standard deviation of time to closed/completed for the test group. Finally, the total number of SMS thresholds exceeded between the time of exceeding the first threshold and the time to closed/completed is calculated for both groups, and the control group is sampled among control carriers with the same distribution of total thresholds exceeded as the test group. Both groups are then followed for 12 months after time to closed/completed and evaluated according to SMS thresholds exceeded.

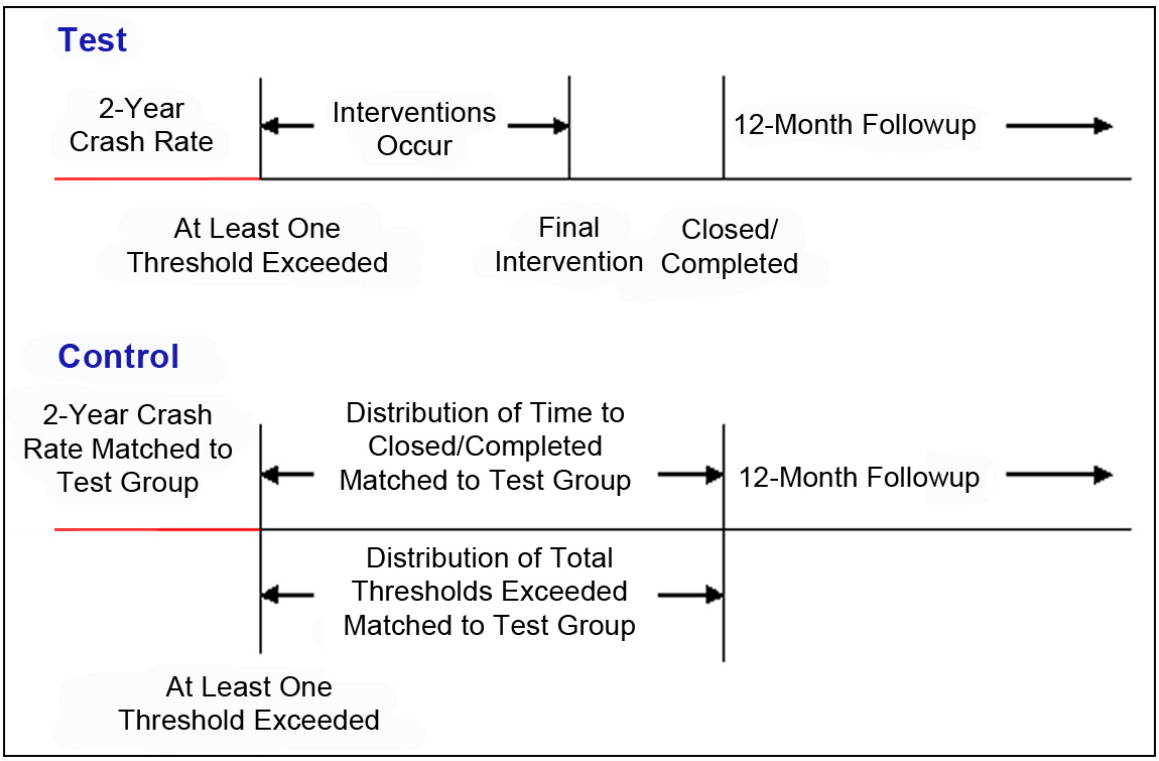

Figure 15. Method for Comparing Test and Control Groups

The procedure to match test and control carriers is designed to ensure that the two groups are as similar as possible in key measures of safety before the evaluation begins, so that any reductions in SMS thresholds exceeded during the 12-month followup period are more likely due to CSA 2010 interventions, and not due to different levels of safety prior to evaluation. The stochastic nature of the sampling procedure is designed to make the two groups as similar as possible on average. This control group is designed to be the most appropriate for comparison to the test group.

In addition to the control group described above, results from two other control groups are presented for comparison. One control group is selected among control carriers that exceeded at least one SMS threshold, but did not receive CRs during the Operational Model Test (February 2008 through June 2010). This group is designed to be a generally large sample of mild carriers from a safety performance point of view. For this control group, the median time to closed/completed for test carriers is added to the month that the first SMS threshold was exceeded to account for the delay in time to closed/completed before the 12-month followup period. The second control group consists of control carriers that did receive CRs after exceeding at least one SMS threshold, and the 12-month followup period begins after CR. This control group is designed to be a generally high-risk group of carriers for comparison to the test group. This control group remains the same for each of the seven intervention patterns considered.

- Several points are worth noting concerning this analysis:
  - For carriers with the intervention pattern of warning letter only, which represents the majority of intervention patterns, it is not necessary to match distributions by time to closed/completed and total thresholds exceeded since the time from exceeding the first SMS threshold to time to closed/completed is generally 1 month. However, crash rates are matched in the 2 years prior to exceeding the first threshold.
  - Since the three variables (crash rate, time to closed/completed, total thresholds exceeded) are necessarily positive, they are analyzed on the natural log scale to make their distributions symmetric and close to normal.
  - It is not necessary for all carriers to have 12 months of followup after the final intervention is closed/completed. For example, a carrier that receives a warning letter a few months before the Operational Model Test ends will likely have only several months of followup. The *cohort* method of analysis used here allows such a carrier to contribute *carrier months* of followup to the study before dropping out. The same is true of control carriers in which time to closed/completed is sampled.
  - *Closed/completed* does not imply that safety issues related to a carrier with intervention(s) have been resolved. It means that the safety investigator has completed certain tasks associated with performing a particular intervention.
  - Except for the warning letter, evaluations for test carriers receiving one investigation (off-site, onsite focused, onsite comprehensive) tend to be based on relatively small sample sizes due to available data during the 12 month followup period. Supporting data for the followup periods for each intervention pattern are provided in Appendix B.

8.1 THE WARNING LETTER

The effectiveness of the warning letter is examined by following test carriers that exceeded at least one SMS threshold and received the warning letter as the only intervention. Since the time to closed/completed for this intervention pattern is generally one month from the time of exceeding the first SMS threshold, matching distributions of time to closed/completed and total thresholds exceeded within that time frame is not considered for the "Control Matched to Test" group (Figure 15). However, test and control groups are matched to have similar distributions of 2-year crash rates just prior to exceeding the first SMS threshold for the matched control group.

Figure 16 shows distributions of log crash rates for the test group and three control groups. In this case, it appears that test carriers are fairly matched with control carriers without CRs since test carriers with warning letter as the only intervention tend to be a mild group in terms of safety. There is one large outlying crash rate in the test group. As expected, control carriers with CRs have the highest crash rates. The control group that is sampled among control carriers to have the same distribution of crash rates as the test carriers is shown at the far right of the figure, and the distributions match closely.

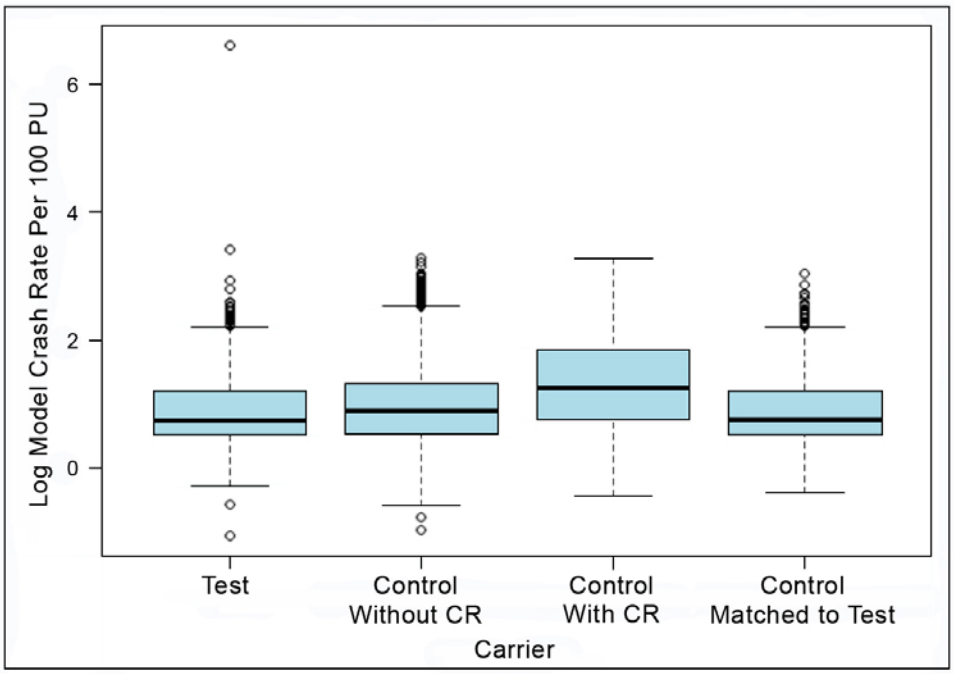

Figure 16. Distributions of Log Crash Rates for Test Group With Warning Letter Only and Three Control Groups

Figure 17 shows the percentage of carriers exceeding at least one SMS threshold for the test group receiving the warning letter only and three control groups (data supporting this figure are provided in Appendix B). This is likely one of the most significant findings in this evaluation. After 12 months of followup, approximately 17 percent of test carriers are exceeding at least one SMS threshold compared to about 45 percent of control carriers matched to test carriers. Table 42 shows that this intervention pattern is by far the most common one, representing about one-third of the total. Since this intervention pattern is the most common, the sample size for the test

group is relatively large: 1,749 test carriers could be identified from the original four States with the warning letter that is closed/completed as the only intervention. Of those carriers, 1,048 had complete data recorded for the 12 month followup period. For analyses of the intervention patterns that follow, sample sizes are not nearly as large.

In this case, the "Control Matched to Test" and the "Control without CR" groups show similar results. Therefore, matching on crash rate did not produce significant differences between the two groups. Figure 17 shows that the distributions of crash rates for these two groups are also somewhat similar, and in this case matching may not have had a big effect. For carriers receiving CRs during the Operational Model Test, almost 60 percent of carriers continue to exceed at least one SMS threshold after 12 months of followup. At time zero, not all carriers were still exceeding at least one threshold. Each group starts at 100 percent to show that all carriers initially exceeded at least one threshold prior to evaluation. However, the first month and the 12 months thereafter reflect the percentage of carriers exceeding at least one threshold for each of the four groups.

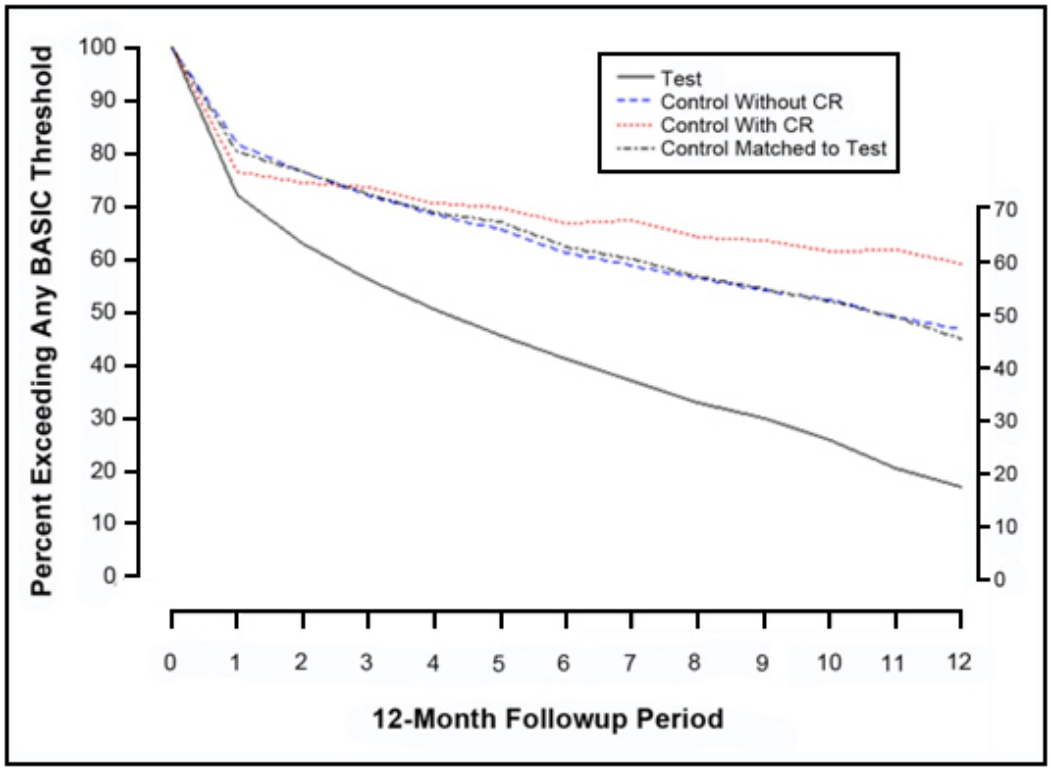

Figure 17. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Warning Letter Only)

8.2 THE OFFSITE INVESTIGATION

The effectiveness of the offsite investigation is examined by following test carriers that exceeded at least one SMS threshold and received the offsite investigation as the only intervention. Again, test carriers are compared with three control groups over 12 months. Since time to closed/completed from the time exceeding the first SMS threshold averaged about 3 or 4 months,

in addition to the crash rate, distributions of time to closed/completed and total thresholds exceeded were matched to the "Control Matched to Test" group. The purpose of Figure 18 is to show that crash rates and total thresholds exceeded have been matched to have similar distributions prior to the 12-month evaluation period for the "Control Matched to Test" control group.

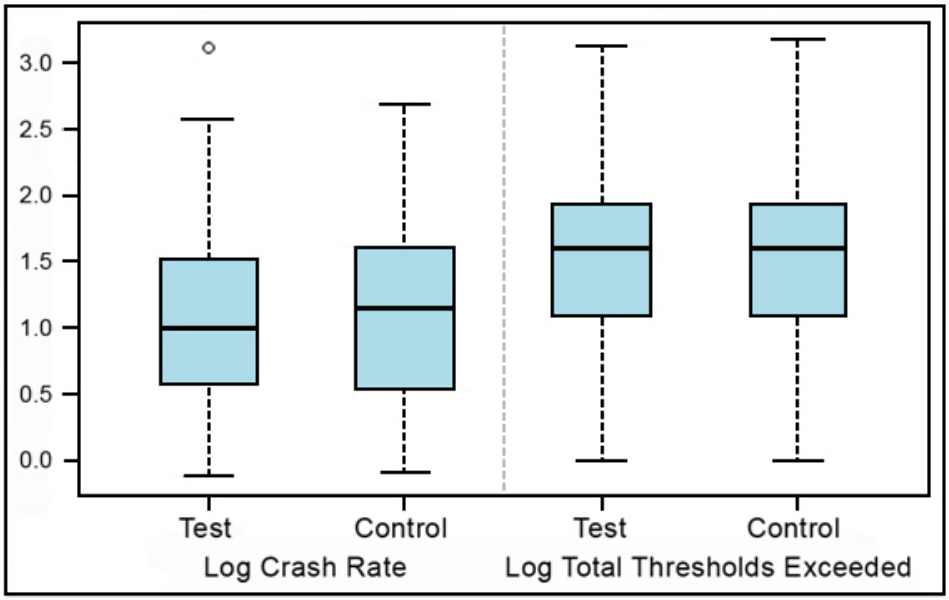

Figure 18. Distributions of Crash Rate and Total Thresholds Exceeded Matching Test and Control Groups (Test—Offsite Investigation Only)

Figure 19 shows distributions of log crash rates for the test group and three control groups. In this case, the test group tends to have a higher crash rate than the control group without CRs, but a lower crash rate than the control group with CRs. The control group that is matched to the test group is the same one shown in Figure 18, and has a distribution of crash rates similar to the test group by design.

For the remaining intervention patterns evaluated, box plots similar to those shown in Figure 18 and Figure 19 are shown in Appendix C. These plots are of distributions matched on crash rates and thresholds exceeded and of crash rates for the test and three control groups.

Figure 20 shows the percentage of carriers exceeding at least one SMS threshold for the test group receiving the offsite investigation only and three control groups. In this plot, the percentage of test carriers exceeding at least one threshold is similar to the control group with CRs for about the first 8 months, then declines to be comparable to the matched control group, and then in the final month declines to about 45 percent. The sample size for the test group is 68 carriers initially, and drops to 49 after 12 months. After 12 months, 54 percent of the "Control Matched to Test" carriers exceed at least one threshold. Results for this control group are intermediate between the two control groups with and without CRs. The sample size for the "Control Matched to Test" control group is approximately 1,000 carriers.

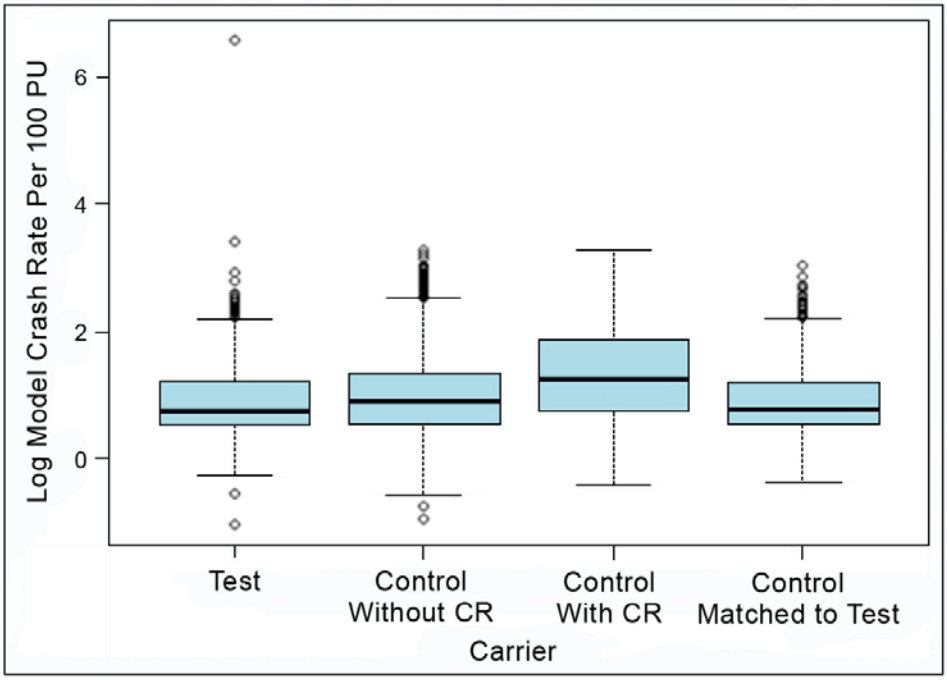

Figure 19. Distributions of Log Crash Rates for Test Group With Offsite Investigation Only and Three Control Groups

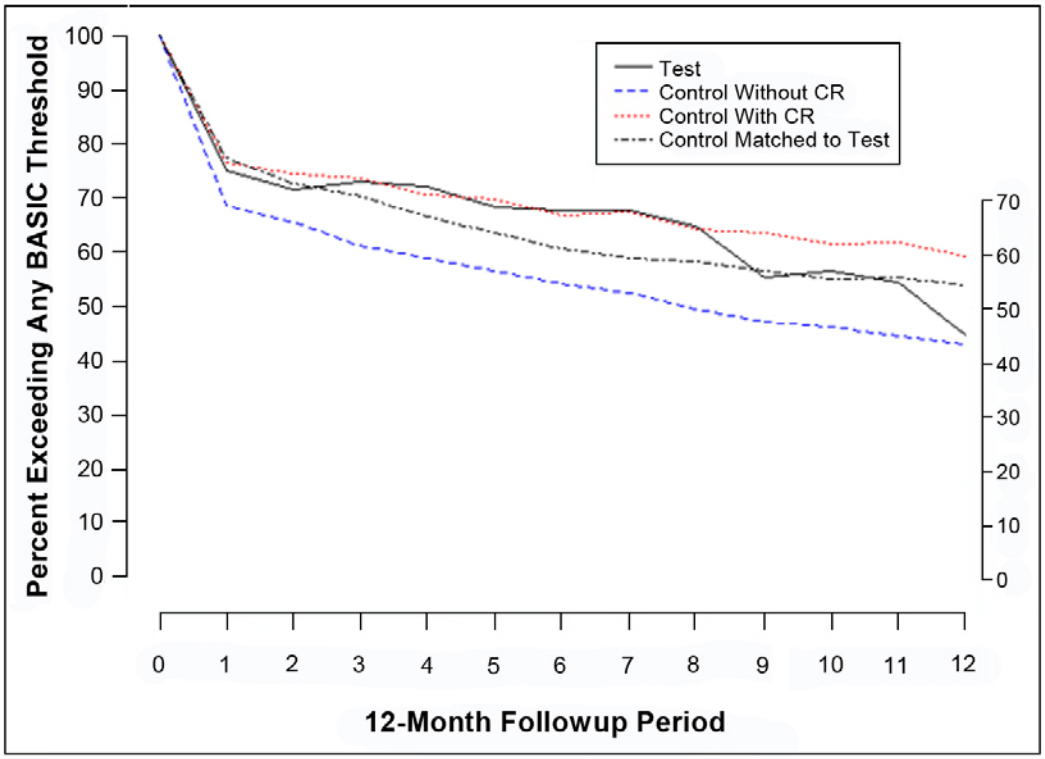

Figure 20. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Offsite Investigation Only)

8.3 THE ONSITE FOCUSED INVESTIGATION

The methods are now applied to the onsite focused investigation. The effectiveness of the onsite focused investigation is examined by following test carriers that exceeded at least one SMS threshold and received the onsite focused investigation as the only intervention. Test carriers are compared with the three control groups over 12 months. The time to closed/completed from the time exceeding the first threshold averaged about 4 or 5 months for the onsite focused investigation. Boxplots showing distributions matched on crash rates and total thresholds exceeded as well as crash rates for the test and three control groups are shown in Appendix C.

Figure 21 shows percentage of any thresholds exceeded for the 12 month followup period after time to intervention closed/completed. There are some similarities and differences between this plot and the one for the offsite investigation (Figure 20). After about 8 or 9 months, the line for the test group crosses lines for both the control group with CRs and the control group matched to the test group. After 12 months the line for the test group reaches 40.6 percent, which is slightly below the 43.1 percent for the control group without CRs. Before 9 months, however, the test group has the largest percent of carriers exceeding at least one SMS threshold. It should be pointed out that the sample size for the test group is 84 carriers at the beginning of the 12 month period and drops to 32 carriers at the end of the period. Supporting data for test and control groups are provided in Appendix B.

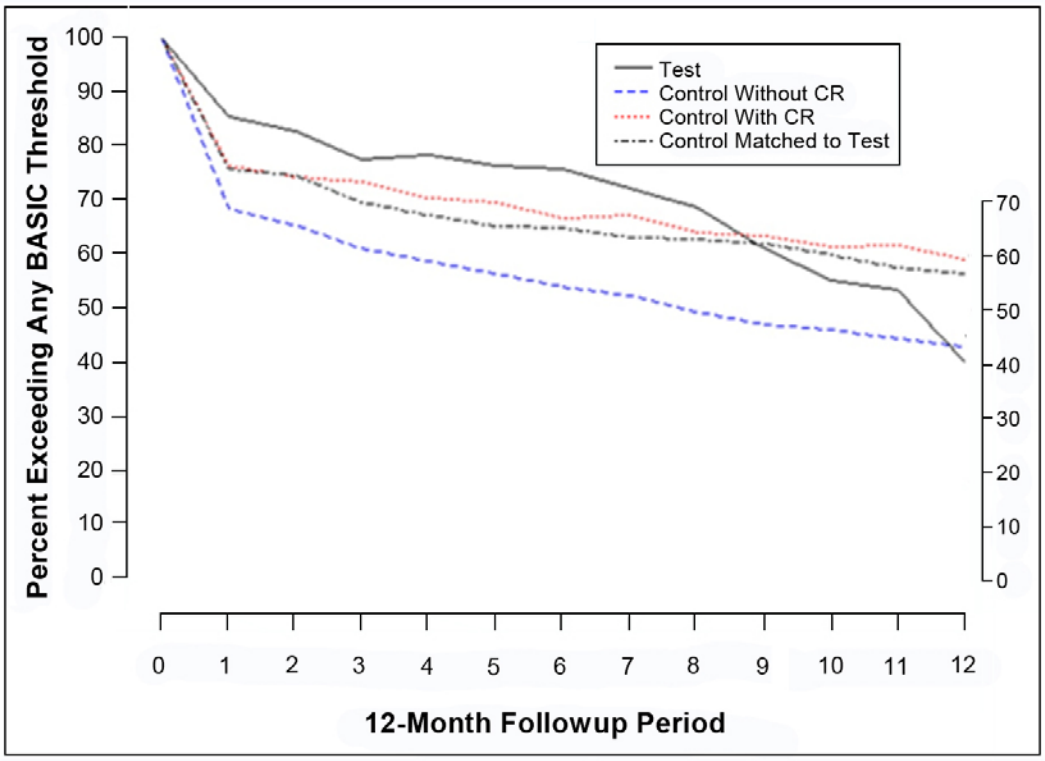

Figure 21. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Onsite Focused Investigation Only)

8.4 THE ONSITE COMPREHENSIVE INVESTIGATION

The effectiveness of the onsite comprehensive investigation is examined. Carriers that exceeded at least one SMS threshold and received the onsite comprehensive investigation as the only intervention are followed over 12 months after the intervention was closed/completed. Test carriers are compared with the three control groups. The time to closed/completed from the time exceeding the first threshold averaged about 6 or 7 months for the onsite comprehensive investigation. Therefore, the time to closed/completed for this intervention is longer than for either the offsite or the onsite focused investigations.

The sample size for evaluating test carriers receiving the onsite comprehensive investigation is small. The number of carriers at the time of closed/completed is 33 and after 12 months the number is 11. This is likely due to the Phase I period conducted in the original four States in which SafeStat A/B carriers were removed from the Operational Model Test.

Figure 22 shows percentage of carriers exceeding at least one SMS threshold over the 12 month period after the onsite comprehensive investigation was closed/completed. In terms of safety, the control group without CR is generally a mild group of carriers and usually provides a lower bound in the plot. The control group with CRs is generally a group of at-risk carriers since some prior behavior prompted a CR, and this group usually provides an upper bound in the plot. The control group matched to the test group is generally intermediate between the other two control groups in the plot since this control has been matched to the test group on several key safety factors (crash rate, time to closed/completed, prior thresholds exceeded). However, in Figure 22, the "Control Matched to Test" group tends to exceed at least one threshold at a higher percentage than the other three groups. After 12 months of followup, about 70 percent of the "Control Matched to Test" carriers still exceed at least one threshold. After 12 months, approximately 46 percent of test carriers continued to exceed at least one threshold, although sample size is small for this group.

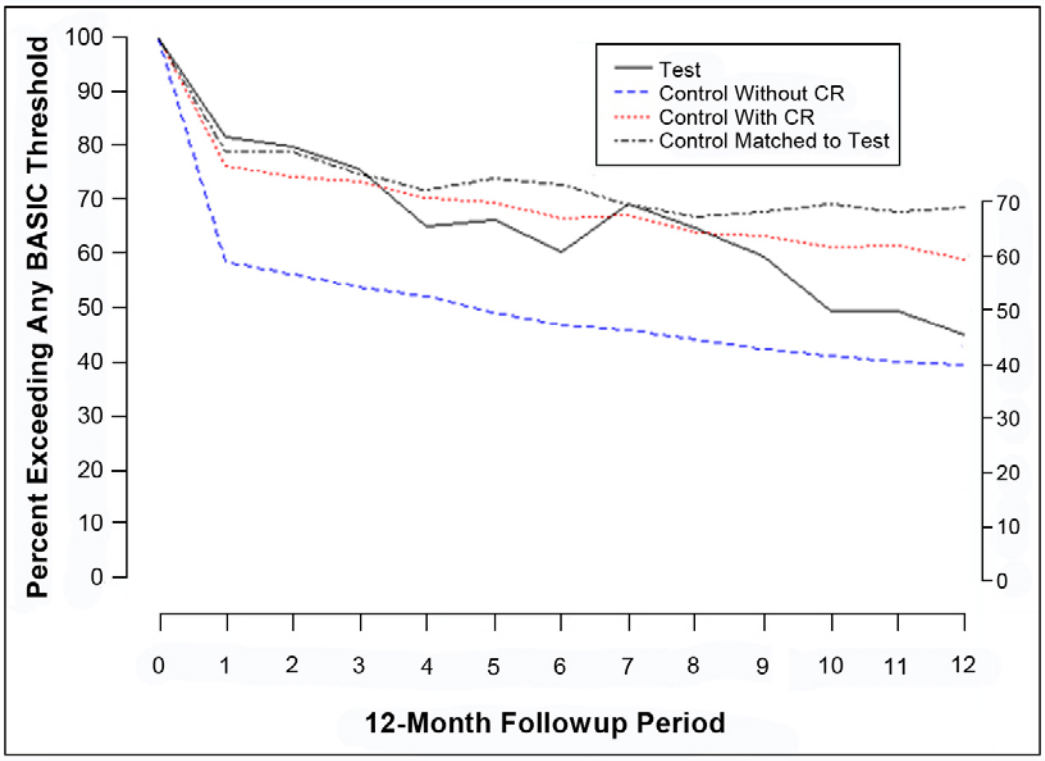

Figure 22. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Onsite Comprehensive Investigation Only)

8.5 THE WARNING LETTER—INVESTIGATION

In this section, carriers receiving a warning letter followed by any of the three investigations (off-site, onsite focused, onsite comprehensive) are considered. The three investigations are combined to maintain adequate sample size for the test group in the 12-month period of evaluation. There were 376 carriers at the time the last intervention was closed/completed, and 130 remaining after 12 months of follow up. For these carriers, the warning letter was escalated to an investigation, so in this case the warning letter may be viewed as not being sufficient at bringing percentile scores under thresholds. The time to final intervention closed/completed was 13 or 14 months for this intervention pattern.

Figure 23 shows the percentage of carriers exceeding at least one threshold over the 12-month period after the last intervention was closed/completed. As in some of the other plots, the control group without CRs gives a lower bound, and the control group with CRs gives an upper bound. As in many of the other plots ending in investigations, after approximately 8 or 9 months the test group starts to show improvement relative to the control groups. After 8 months, about 50 percent of the test group and about 50 percent of the control group matched to the test group exceed at least one threshold. After that time, the percentage for the test group is lower, although the difference is modest. After 12 months, the percentage for the test group is 41.5 percent and for the matched control group it is 46.2 percent.

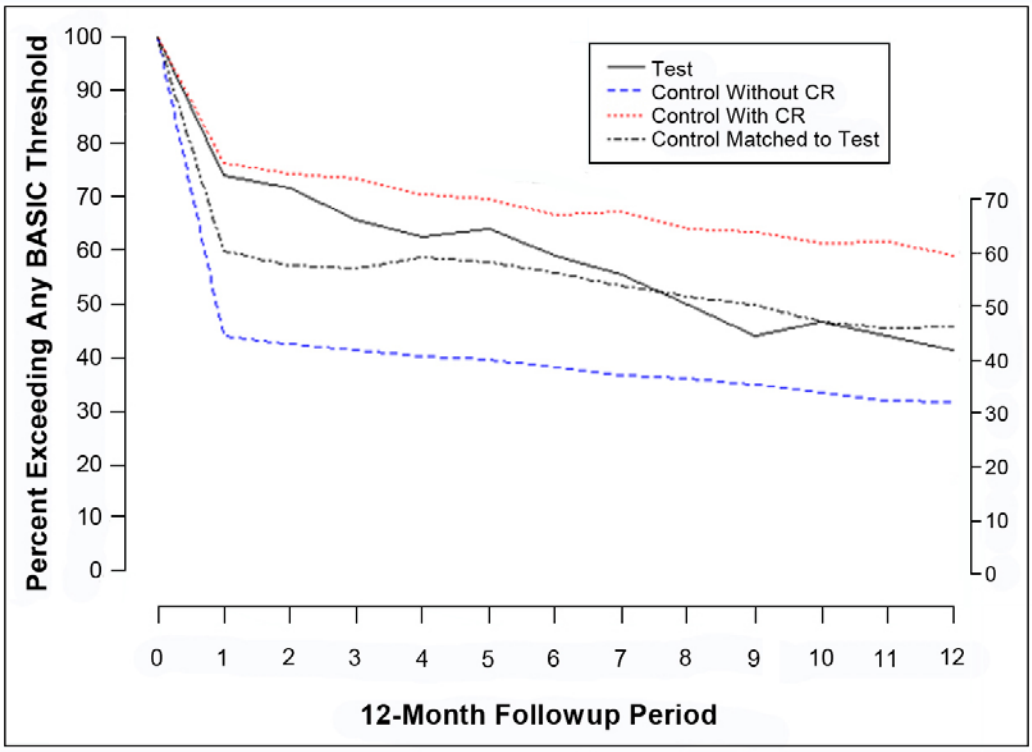

Figure 23. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test—Warning Letter and Investigation)

8.6 THE WARNING LETTER—INVESTIGATION—CSP

The intervention pattern beginning with a warning letter, followed by one of the investigations (offsite, onsite focused, onsite comprehensive), and ending with a CSP is considered. The sample size for the test group in which the CSP was closed/completed is fairly large. There are 385 carriers at the time the CSP was closed/completed and 185 after 12 months of followup. The average time to closed/completed from the time of first exceeding any SMS threshold is approximately 11 or 12 months. Therefore, on average it took about 1 year from the time of exceeding the first threshold until the CSP was closed/completed.

Figure 24 shows percentage of carriers exceeding at least one threshold for the test group and three control groups over the 12-month period. The control without CR provides a lower bound while the control with CR provides an upper bound. The control matched to the test group is intermediate between the two. The test group shows results similar to the control group with CR for approximately 6 months, and then improves. After 12 months, however, the percentage of carriers exceeding at least one BASIC threshold drops to approximately the same level as the matched control group (approximately 50 percent of carriers).

Carriers with multiple interventions take longer until time to closed/completed. It may be that 12 months is not a sufficient period of followup to detect differences for carriers with intervention patterns involving multiple interventions over a generally long timeframe. Even though the CSP is closed/completed prior to the 12-month followup, carriers are likely being monitored well after

that time. Note that carriers receiving one intervention that was an investigation, for example, the offsite, the onsite focused, or the onsite comprehensive, tended to improve after the 8-month lag time after the investigation was closed/completed. For intervention patterns involving multiple interventions, the lag time may be longer.

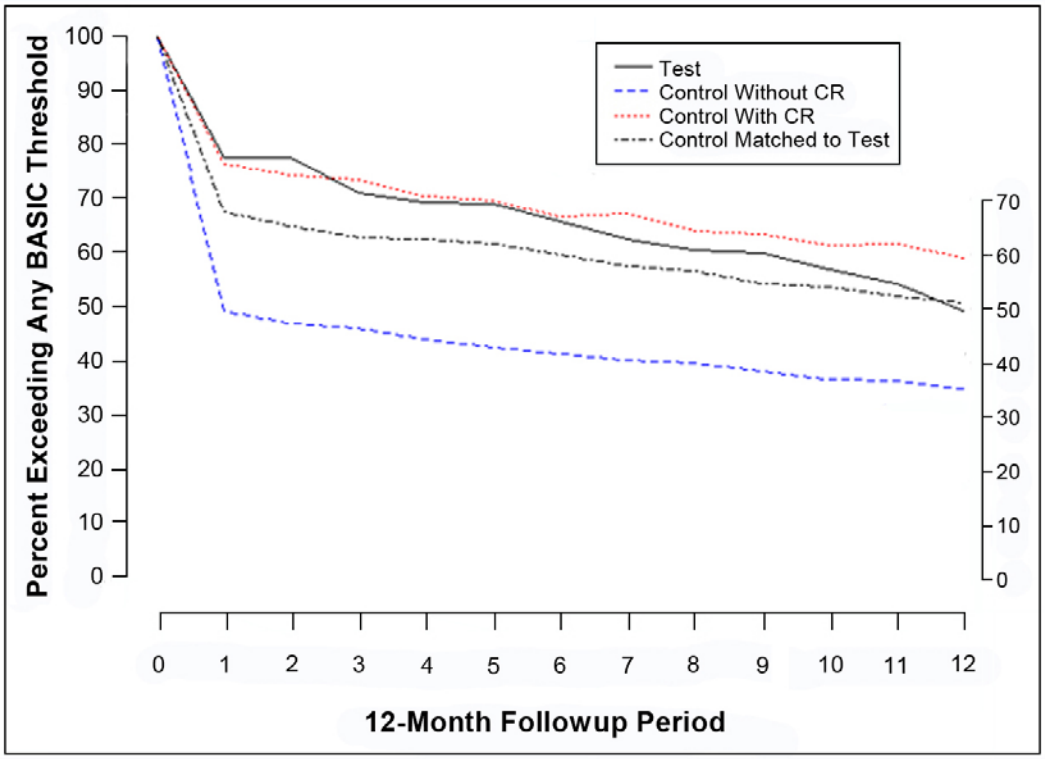

Figure 24. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test: Warning Letter—Investigation—CSP)

8.7 INTERVENTIONS ENDING IN NOC

The last intervention pattern considered is any pattern that ended with a NOC. The average time to closed/completed for this pattern was between 11 and 12 months. The sample size for the test group is 267 carriers at the time of closed/completed and 95 after 12 months of followup.

It makes sense that carriers that require the most severe intervention patterns have the hardest time coming into line. Carriers with interventions ending with the NOC clearly fall into this category. In Figure 25, percentages for the matched controls are similar to those for carriers with CRs. The test group has the highest percentages until the seven month mark, at which time it improves relative to the controls with CRs and the controls matched to the test carriers. At the 12-month mark, the percentage of carriers exceeding at least one BASIC threshold is 57.9 for the test group carriers and 60.7 for the matched control group.

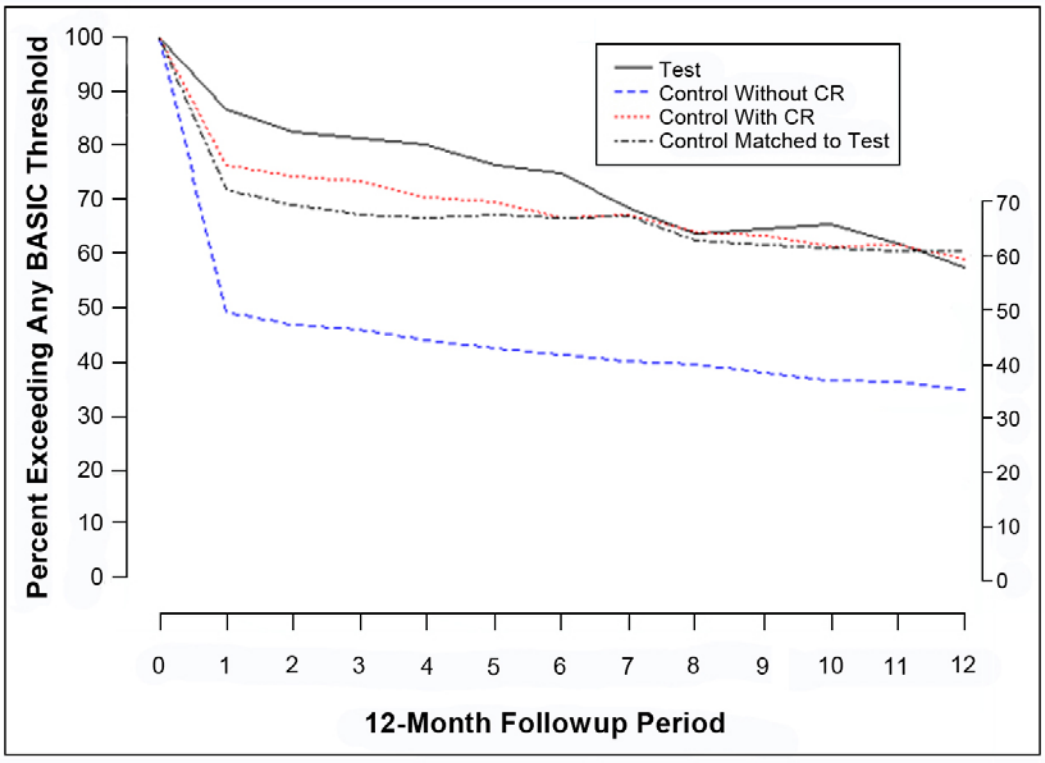

Figure 25. Percent of Carriers Exceeding At Least One SMS Threshold During 12-Month Followup (Test: Ending in NOC)

[This page intentionally left blank.]

9. SAFETY COMPARISON OF SAFESTAT A/B CARRIERS

The SafeStat ranking of carriers is one of the current criteria used by FMCSA to prioritize carriers for CRs. A question of interest is how test and control SafeStat A/B carriers fared in terms of safety performance during the Operational Model Test. One way to answer this question is to compare the two groups with respect to certain safety metrics over a designated period of time. Four metrics that are particularly relevant to carriers that have been classified as SafeStat A/B were used to evaluate safety performance:

- Percent carriers exceeding at least two BASIC thresholds.
- Percent carriers exceeding the Unsafe Driving or Crash Indicator thresholds.
- Average number of BASIC thresholds exceeded.
- Crash rate predicted by the model described in Section 5.8.

Carriers identified as SafeStat A/B often exceed more than one BASIC threshold, so it seems reasonable that the first metric evaluates carriers exceeding at least two thresholds. The Unsafe Driving and Crash Indicator measures were shown to be most related to crash rates, so the second metric evaluates carriers exceeding either one or both of those thresholds. The third metric evaluates carriers on total number of BASIC thresholds exceeded, and the final metric is the crash rate as predicted by the model described in Section 5.8.

Not all test group A/B carriers received interventions under CSA 2010 and not all control group A/B carriers received CRs under the current system. However, as has been shown, CSA 2010 touches more carriers with interventions than does the current system with CRs. Analysis of the data shows that approximately 95 percent of test carriers identified as SafeStat A/B at some time during the Operational Model Test received interventions under the CSA 2010 system, and they were generally high level investigations (onsite focused, onsite comprehensive), or CSPs or NOCs.. Approximately 70 percent of control carriers identified as SafeStat A/B received CRs during the same time period under the current system. To answer the original question of how test and control SafeStat A/B carriers fared under the two systems, results are restricted to the 95 percent of test carriers that received interventions under the new system (CSA 2010), and the 70 percent of control carriers that received CRs under the current system.

It is not necessary to match test and control carriers in terms of safety performance prior to the evaluation, as was done in section 8, because SafeStat A/B carriers have already been identified as those with specific safety performance issues and the two groups are well-matched. Sample sizes are large enough to evaluate both test and control carriers over an 18 month time period. Supporting data in Appendix D shows that there are approximately 800 test A/B carriers for evaluation at time zero, and about 250 after eighteen months. The corresponding numbers for control A/B carriers are 600 at time zero and 150 after eighteen months. Time zero corresponds to the month that the carrier was first identified as a SafeStat A/B carrier.

Figure 26 shows the comparison between test and control SafeStat A/B carriers according to the four safety performance measures. The test group carriers perform better in each case. In two of the four plots (percent exceeding at least two BASIC thresholds, average number of BASIC

thresholds exceeded), the lines for the test and control groups are similar for a number of months and then separate with improved performance for the test group in the later months. For the other two performance measures (percent exceeding the Unsafe Driving or the Crash Indicator BASICs, average predicted log crash rate), the test group continuously performs better than the control group from beginning to end. Crash rates predicted by the model show small improvement over time, but they are shown on the log scale which is a common practice when presenting rates. Log crash rates for the test group improve over time relative to the control group. It should be noted that in this analysis, all A/B test carriers received CSA 2010 interventions and all A/B control carriers received CR's at some time during the 29 months of the Operational Model Test. However, no distinction was made as to when interventions or CRs occurred. They may have occurred either before or after the month the carrier was first identified as a SafeStat A/B carrier. Nevertheless, the test group shows better performance than the control group in each of the four plots.

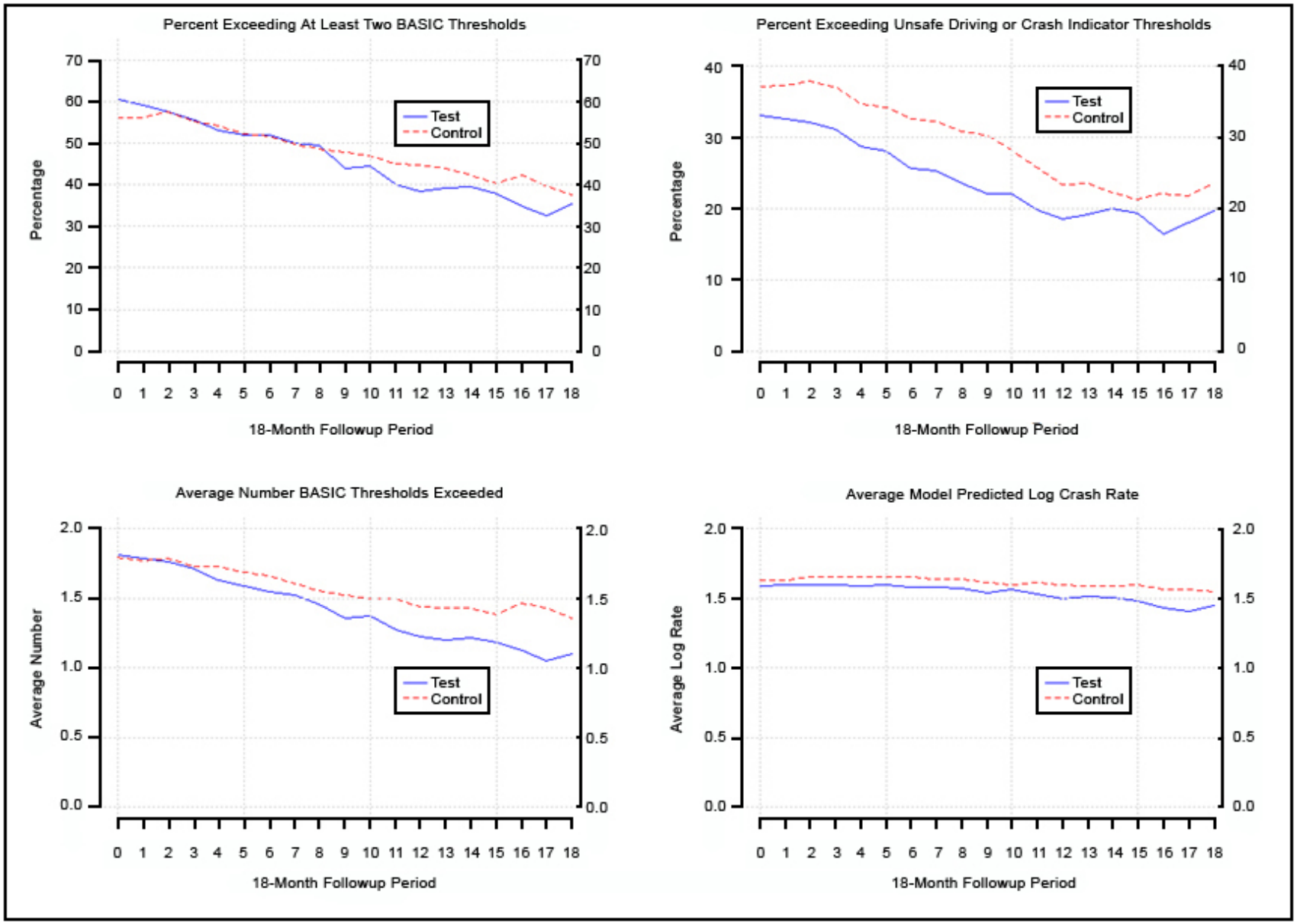

Figure 26. Comparison of Test and Control SafeStat A/B Carriers on Four Safety Performance Measures During the Operational Model Test

[This page intentionally left blank.]

10. COSTS TO FMCSA OF CSA 2010 INTERVENTIONS AND CRs, AND TOTAL AMOUNTS CLAIMED

Two of the major goals of the CSA 2010 Operational Model Test are to have more contact with carriers (more touches) and to use Agency resources more effectively. Using Agency resources more effectively is largely related to performing cost-effective interventions according to current Agency staffing levels. In this section, costs to FMCSA of performing CSA 2010 interventions on carriers in the Operational Model Test are investigated, as well as the dollar amounts claimed. Results pertain to the original four States, namely, Colorado, Georgia, Missouri, and New Jersey, and are based on the 8 months of data collected during the Phase II period from October 2008 through May 2009 inclusive. The analysis is based on all interventions for test group carriers that were classified as closed/completed (all tasks for the intervention completed and the intervention marked closed). For comparison, costs to the Agency of performing CRs to control group carriers under the current enforcement program are also calculated.

Staff personnel record four variables relevant to calculating costs that are uploaded into data files when conducting CSA 2010 interventions: labor hours, Government miles, vouchers, and other expenses. For test group carriers receiving CSA 2010 interventions, the formula used for calculating cost in dollars is: Cost equals (adjusted hourly rate times labor hours) plus (0.585 times Government travel miles) plus vouchers plus other expenses.

The adjusted hourly rate is the basic hourly rate according to the General Schedule (GS) salary table effective January 2009 plus 32.85 percent fringe benefits and 12 percent overhead applied to the basic hourly rate plus fringe benefits. Adjusted hourly rates for staff personnel responsible for tasks when conducting interventions are calculated for four labor categories and are shown in Table 44 below:

Table 44. Adjusted Hourly Rates for CSA 2010 Staff Personnel

Category	Grade	Basic Hourly Rate	Fringe Benefits	Overhead	Adjusted Hourly Rate
Intervention Manager	GS-13 Step 5	$38.35	$12.60	$6.11	$57.06
Safety Investigator	GS-12 Step 5	$32.25	$10.59	$5.14	$47.99
Program Analyst	GS-9 Step 1	$19.62	$6.45	$3.13	$29.19
Clerical Specialist	GS-5 Step 1	$12.95	$4.25	$2.06	$19.27

The quantity 0.585 in the cost formula is the amount used by FMCSA to cost Government travel miles in dollars. For control group carriers receiving CRs, it is assumed that the safety investigator performed all of the labor, and the adjusted hourly rate of $47.99 is applied to all labor hours.

Distributions of cost and other related variables presented tend to be skewed due to the nature of the data and some large outliers. Therefore, along with the average, the median, minimum, and maximum values are presented to provide an indication of the effect of outliers on the resulting estimates.

Since the cost of a warning letter is nominal, test group carriers were evaluated to estimate the costs associated with all interventions except warning letters. Overall, 603 test group carriers with 940 interventions were identified in the Phase II period between October 2008, and May 2009. All interventions are closed/completed. Of the 603 test carriers, 17 did not have data recorded for labor hours and other relevant expenses. Table 45 shows totals for labor hours, Government travel miles, vouchers, expenses, and cost for the 586 carriers with 920 interventions for which complete data are recorded. The safety investigator accounts for a large percentage of the totals. The total cost in dollars during the 8-month period for which data are complete is approximately $442,000.

Table 45. Total Cost in Dollars by Labor Category for All Interventions (October 2008–May 2009, Phase II)

Labor Category	Labor Hours	Government Miles	Vouchers	Expenses	Cost
Intervention Manager	37.25	0.00	0.00	0.00	$2,125.49
Safety Investigator	8,680.75	23,402.00	5,354.10	1,056.34	$436,689.80
Program Analyst	89.75	0.00	0.00	56.78	$2,676.58
Clerical Specialist	9.25	0.00	0.00	0.00	$178.25
Total	**8,817.00**	**23,402.00**	**5,354.10**	**1,113.12**	**$441,670.12**

Table 46 shows statistics of the five variables per carrier. The values in the average column can be calculated by dividing the totals in Table 45 above by 586. The median, minimum, and maximum provide an indication of the effect of outliers on the average. The estimated average cost per carrier for all interventions performed is $754.00, while the median cost is $590.00.

Table 46. Cost and Related Statistics Per Carrier for All Interventions Performed (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	586	$753.70	$590.03	$7.30	$7,838.72
Labor Hours	586	15.05	11.75	0.25	153.50
Government Miles	586	39.94	0.00	0.00	779.00
Vouchers	586	9.14	0.00	0.00	424.00
Expenses	586	1.90	0.00	0.00	337.87

In order to adjust total values to account for the 17 carriers that received interventions, but for which labor data were not recorded, the average cost of each intervention per carrier (see Table 52 through Table 57 which estimate costs by intervention type) was applied to the 17 carriers with missing labor data. The 17 carriers with missing labor data had 20 interventions as shown in Table 47 below. Eleven of the 20 are CSPs. Among the interventions other than the warning letter, the CSP is the least expensive to conduct. A reasonable estimate for the total cost of all interventions is about $450,000. Multiplying the 8-month total by 1.5 gives an estimated annual cost of $675,000.

Table 47. Adjusted Total Cost in Dollars for All Interventions Performed by Applying Average Cost of Each Intervention (October 2008–May 2009, Phase II)

Intervention	N	Average Cost	Total Cost
CSA	11	$94.78	$1,042.58
NOV	0	$117.87	$0.00
NOC	3	$427.56	$1,282.68
Offsite Investigation	0	$450.70	$0.00
Onsite Focused Investigation	5	$677.06	$3,385.30
Onsite Comprehensive Investigation	1	$1,038.50	$1,038.50
Total for Missing Data	20	–	$6,749.06
Total Cost from Table 45			$441,670.12
Adjusted Total Cost			$448,419.18

A similar analysis as the one presented above is performed, restricting test group carriers to those classified as SafeStat A/B during the same Phase II period with interventions that are closed/completed. Once again, values are broken down by labor category and totals are presented in the bottom row. Overall, 174 test group carriers with 302 interventions were identified in the Phase II period from October 2008 through May 2009. Of the 174 test carriers, 4 do not have data recorded for labor hours and other relevant expenses. Table 48 shows totals for labor hours, government travel miles, vouchers, expenses, and cost for the 170 carriers with 297 interventions for which complete data are recorded. The total cost in dollars is approximately $175,000.

Table 48. Total Cost in Dollars by Labor Category for All Interventions Limited to SafeStat A/B Carriers (October 2008 through May 2009, Phase II)

Labor Category	Labor Hours	Government Miles	Vouchers	Expenses	Cost
Intervention Manager	10.75	0.00	0.00	0.00	$613.40
Safety Investigator	3,486.75	9,549.00	1,505.19	104.48	$174,524.97
Program Analyst	12.50	0.00	0.00	6.48	$371.36
Clerical Specialist	0.00	0.00	0.00	0.00	$0.00
Total	**3,510.00**	**9,549.00**	**1,505.19**	**110.96**	**$175,509.72**

Table 49 shows statistics of the five variables per carrier. The values in the average column can be calculated by dividing the totals in Table 48 above by 170. The estimated average cost per SafeStat A/B carrier for interventions performed is $1,032, while the median cost is $824.

Table 49. Cost and Related Statistics Per Carrier for All Interventions Performed Limited to SafeStat A/B Carriers (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	170	$1,032.41	$824.14	$7.30	$6,046.74
Labor Hours	170	20.65	17.00	0.25	126.00
Government Miles	170	56.17	0.00	0.00	779.00
Vouchers	170	8.85	0.00	0.00	360.00
Expenses	170	0.65	0.00	0.00	60.00

In this case, only four carriers received interventions for which no labor data were recorded. However, to be consistent, adjusted totals are presented. Table 50 shows results which represent adjusted totals for all 174 test group carriers that were classified as SafeStat A/B and received interventions during the Phase II period. The adjusted total cost is about $180,000.

Table 50. Adjusted Total Cost in Dollars for All Interventions Performed Limited to SafeStat A/B Carriers (October 2008–May 2009, Phase II)

Totals	Labor Hours	Government Miles	Vouchers	Expenses	Cost
Totals from Table 48	3,510.00	9,549.00	1,505.19	110.96	$175,509.72
Averages from Table 49	20.65	56.17	8.85	0.65	$1,032.41
Averages x 4	82.59	224.68	35.42	2.61	$4,129.64
Adjusted Totals	3,592.59	9,773.68	1,540.61	113.57	$179,639.36

For purposes of comparison, costs are now calculated for non-test carriers receiving CRs during the same Phase II period used to evaluate test group carriers. Non-test carriers are those classified as control, removed, excluded, or nonparticipating carriers (all but test carriers) in the original four States. Table 51 shows carrier level statistics for non-test carriers receiving CRs as recorded by safety investigators. For non-test carriers, it is not possible to present totals by labor category since data are not recorded at that level. Therefore, the adjusted hourly rate of a safety investigator is used to calculate cost. Overall, 364 non-test carriers with 481 CRs were identified during the Phase II period from October 2008, through May 2009. The average cost per carrier is $1,438, while the median cost is $1,058. No adjustment for missing data is required here since all 364 carriers have labor data recorded.

Table 51. Cost and Related Statistics Per Carrier for All CRs Performed for Non-Test Carriers (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	364	$1,438.02	$1,058.43	$191.65	$16,935.63
Labor Hours	364	27.30	19.84	3.25	345.00
Government Miles	364	170.59	120.00	0.00	1,300.00
Vouchers	364	27.06	0.00	0.00	2,112.00
Expenses	364	0.87	0.00	0.00	93.00

Tables analogous to the ones presented above for test group carriers receiving CSA 2010 interventions are now presented for each of the specific intervention types, excluding the warning letter. The interventions considered are CSP, NOV, and NOC. Results for offsite and onsite focused and comprehensive investigations are also presented. All interventions are closed/completed. Note that the number of carriers will not add up to 586 since many of these carriers received more than one intervention and can be counted multiple times in the various tables. The intervention types are presented in order of cost from lowest to highest.

Of the intervention types, the CSP is the least expensive to perform. Table 52 shows that among 244 carriers evaluated, the average cost is about $95 and the median cost is $72.

Table 52. Cost and Related Statistics Per Carrier for the CSP (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	244	$94.78	$71.99	$7.30	$551.89
Labor Hours	244	1.97	1.50	0.25	11.50
Government Miles	244	0.45	0.00	0.00	90.00
Vouchers	244	0.00	0.00	0.00	0.00
Expenses	244	0.05	0.00	0.00	6.07

The NOV is the rarest of all interventions. In the 8-month period of Phase II under consideration, only six carriers received NOVs. Table 53 shows that the average cost of the NOV is approximately $118 and the median cost is $96.

Table 53. Cost and Related Statistics Per Carrier for the NOV (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	6	$117.87	$95.98	$21.08	$254.22
Labor Hours	6	2.46	2.00	0.50	5.25
Government Miles	6	0.00	0.00	0.00	0.00
Vouchers	6	0.00	0.00	0.00	0.00
Expenses	6	1.08	0.00	0.00	6.48

The NOC was performed on 115 carriers during the 8 months under consideration. Table 54 shows that the average cost per carrier is $428 and the median cost is $192. The disparity between the average and the median is due to a few large values that skew the average to the right. For example, note the maximum value of $4,799 is an extreme case. For this reason, both the average and the median are presented.

Table 54. Cost and Related Statistics Per Carrier for the NOC (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	115	$427.56	$191.96	$7.30	$4,799.00
Labor Hours	115	8.90	4.00	0.25	100.00
Government Miles	115	0.56	0.00	0.00	64.00
Vouchers	115	0.00	0.00	0.00	0.00
Expenses	115	0.00	0.00	0.00	0.00

Cost-related results for the three investigations are presented. The investigations tend to be more costly than the CSP, NOV, or NOC. Table 55 shows that for the 149 carriers receiving the offsite investigation, the average cost is $450 and the median cost $406.

Table 55. Cost and Related Statistics Per Carrier for the Offsite Investigation (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	149	$450.70	$405.94	$14.60	$1,435.00
Labor Hours	149	9.55	8.50	0.50	30.00
Government Miles	149	0.77	0.00	0.00	70.00
Vouchers	149	0.27	0.00	0.00	40.96
Expenses	149	0.75	0.00	0.00	14.49

Table 56 shows cost-related results for 249 carriers receiving the onsite focused investigation. The average cost per carrier for this intervention is approximately $677 and the median cost is $588.

Table 56. Cost and Related Statistics Per Carrier for the Onsite Focused Investigation (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	249	$677.06	$587.88	$7.30	$2,207.54
Labor Hours	249	13.31	11.75	0.25	46.00
Government Miles	249	48.78	0.00	0.00	400.00
Vouchers	249	8.55	0.00	0.00	360.00
Expenses	249	2.08	0.00	0.00	337.87

Finally, Table 57 shows cost-related results for the 128 carriers that received the onsite comprehensive investigation. This investigation is similar to a full CR and the average cost per carrier is approximately $1,040 while the median cost is about $880.

Table 57. Cost and Related Statistics Per Carrier for the Onsite Comprehensive Investigation (October 2008–May 2009, Phase II)

Statistic	Carriers	Average	Median	Minimum	Maximum
Cost	128	$1,038.50	$876.84	$7.30	$7,736.67
Labor hours	128	20.01	16.88	0.25	151.50
Government miles	128	85.68	10.00	0.00	779.00
Vouchers	128	24.88	0.00	0.00	424.00
Expenses	128	3.62	0.00	0.00	229.60

Table 58 provides a summary of the estimated costs associated with each CSA intervention type and the CR. Estimated annual costs for the test and non-test groups are also shown. These costs pertain to carriers in the original four States.

Table 58. Summary of Costs in Dollars for Test Group Interventions and Non-Test Group CRs

CSA Intervention Types	Average Cost	Median
CSP	$95	$72
NOV	$118	$96
NOC	$428	$192
Offsite Investigation	$451	$406
Onsite Focused Investigation	$677	$588
Onsite Comprehensive Investigation	$1,038	$877
Warning Letter	Nominal	Nominal
Estimated Annual Test Group Costs		$675,000
Non-Test Group		
CR	$1,438	$1,058
Estimated Annual Non-Test Group Costs		$785,000

In addition to the costs to FMCSA of performing interventions on carriers in the CSA 2010 Operational Model Test, dollar amounts claimed from test and control group carriers are also presented. Dollar amounts settled are essentially the same as dollar amounts claimed and are not shown because any differences are negligible. The amounts shown are those claimed during the 29 months of the Operational Model Test evaluation period and are restricted to those carriers in the original four states. Table 59 shows number of carriers, number of claims, and total dollar amounts claimed for both test and control carriers. Most carriers received one claim, but some received two claims, and a few received three, which explains why the number of claims is greater than the number of carriers. In addition to the total amount claimed, the average and median are shown. Because a few claims were large, the average is skewed as a large number and does not reflect the *typical* claim. The median is presented to represent the typical claim.

Table 59. Dollar Amounts Claimed From Test and Control Carriers (Original Four States, 29 Months)

Carrier Group	Number of Carriers With Claims	Number of Claims	Total Amount Claimed ($)	Average Per Claim	Median Claim
Test	658	720	$3,611,547	$5,016	$2,200
Control	601	640	$4,029,516	$6,296	$2,480
Total	**1,259**	**1,360**	**$7,641,063**	**$5,618**	**$2,265**

Based on the results in the table, it appears that there are no significant differences in the experience between test and control groups with respect to claims made. The number of carriers against whom claims were made is about the same. The average per claim is somewhat higher for the control group, but results are skewed by a few very large claims. Because of this, the median is a better measure of comparison. The median claim for the control group is slightly higher.

11. QUALITATIVE ANALYSIS (PROCESS REVIEW) AND DESCRIPTIVE STATISTICS

A survey of the field staff implementing the CSA program was conducted. The purpose was to collect their insights and perspective on both the process and results of the new system of identifying unsafe carriers and intervening. The survey was reasonably short in terms of the number of questions, but covered the broad areas of implementing the interventions, collecting and processing documentation from the carriers, and perspectives on the effectiveness of different interventions. The specific areas for the qualitative evaluation specified in the original statement of work for this project are the following:

- Which interventions were effective in terms of ease of implementation?
- To what extent was requesting documentation from carriers a problem?
- To what extent was processing requested documentation from carriers a problem?
- Which interventions should be made a part of a new FMCSA enforcement model, which should not?

In implementing this survey in practice, a short survey instrument was developed consisting of eight primary questions with four additional questions probing for detail. The complete survey instrument is included in Appendix F. The questions all ask for a discussion response. None of the questions asked for a numerical rating or a categorical response, in which respondents would have to choose from a fixed list of answers. All were discussion questions, which allow respondents more latitude.

The survey instrument was sent to 18 respondents—only nine of which were non-Federal (i.e., State) employees—with at least one from each of the original four test States, as well as representatives from all the other States that were added as "100 percent" test States. All respondents were assured of confidentiality, to get as frank responses as possible. The respondents were told that, within a State, they could collaborate on the answers if they wished. We received responses from eight States, including all four of the original test States: New Jersey, Colorado, Georgia, and Missouri.

Overall, the evaluation of the CSA was quite positive, both with regard to the process and procedures as well as the outcomes in terms of identifying the right carriers for interventions and having the right tools to improve carrier behavior. There were some critiques of the method of prioritizing carriers for interventions and questions about the quality of the data used to score the BASICs (mostly having to do with the timeliness of the data) but overall the people who were actually operating the new system believe that the new CSA model represents a significant improvement over the prior SafeStat system.

In terms of the CSA 2010 process for prioritizing carriers for intervention, one respondent remarked that he was seeing more carriers than in the past, and a different group of carriers. "I [attribute] the change to the new SMS safety yardstick that is being applied more equally across all carriers." Thus, more carriers are being identified for attention, and they were seeing many carriers that they had not seen in the past. This respondent said that they were still seeing the

"traditional problem carriers" that they frequently dealt with but also carriers with which FMCSA had not had contact in the past.

Most of the respondents felt that the prioritization system, to identify and rank carriers for action works well, though there were some questions about the balance struck in the new system between the number of BASICs failed by carriers and the length of time since the last contact. The investigation priority rules combine both the number of failed BASICs and the time since the last intervention. Carriers with one or more BASIC thresholds exceeded but which have not had a contact within the past 2 years are given a higher priority over carriers with more failed BASICs but who have been contacted (intervened upon) more recently. One respondent asked, "If a company is high risk and our own roadside information confirms this why should I assign a two BASIC carrier [two failed BASICs][for investigation] prior to a five BASIC carrier [five failed BASICs] in investigate two [second highest priority]?" Others shared this concern, though all acknowledged that some sort of trade-off is necessary between the length of time since the last contact and the number of failed BASICs.

In related observations, several respondents commented on the effect of the amount of data from inspections and the time lag between the data (e.g., inspection results) and carriers' current deficiencies. One respondent stated that older data carries too much weight and "we're often left visiting carriers with few or no violations in the past year. . . . If they don't have issues NOW, we're . . . wasting time and energy seeing carriers that don't necessarily need to be seen." Another respondent urged that FMCSA "ensure adequate roadside inspection activity in the past 6 to 12 months prior to assignment," to provide the most current picture of carrier safety possible. These observations parallel the complaint above about the prioritization of investigations, that carriers who had not been contacted by FMCSA for a long period were prioritized over carriers with more failed BASICs but who had been seen more recently. The field staff is clearly focused on working with the highest risk carriers based on the most recent and complete data possible.

Respondents were also asked about the intervention types recommended for carriers. Most indicated that the recommended intervention types were appropriate and that each intervention type was useful in different situations, depending on the response of the carrier. "We are in support of all established interventions. Each have their place. The offsite [investigation] is more specialized but can be effective." Another commented that " the recommended types of interventions shown in CSI [Comprehensive Safety Information Web site] for the interstate carriers have been spot on." A couple commented that they appreciate and use flexibility in selecting intervention types. The general sentiment was that the intervention types recommended in CSI were appropriate, that each intervention type has a role to play, but that the investigation managers exercised their judgment in applying interventions in specific cases, based on circumstances. "Even the cutoff limit for what falls into full [onsite comprehensive investigation] or focused [onsite investigation] is normally not an issues [*sic*] since once a company gets beyond a certain point the comfort level of investigation manager/safety investigator] dictates a full [review] regardless." In other words, while the investigation managers typically find the recommended interventions appropriate, they exercised discretion, recognizing the reality that the boundaries between different intervention types are flexible in some instances.

Several respondents indicated that they sometimes override the recommended prioritization of a carrier for intervention. This might mean investigating a carrier with a priority of three over one with two investigation managers in the test States indicated that this was done at times, but primarily "due to resource placement during the test phase." During the Operational Model Test, in some cases safety investigators were assigned to either the test or control group within a State, meaning they either dealt with test carriers or with control carriers. Depending on their base location and the location of a carrier, situations might occur requiring a safety investigator to travel a great distance to deal with a problem carrier. investigation managers in most of the test States indicated that sometimes they assign safety investigators to investigate nearby carriers even if they were slightly lower in priority, simply to allocate resources more efficiently. One respondent also indicated that he regularly overrode recommended offsite investigations when it would be quicker to simply drop by the carrier's office to get the information needed.

Respondents also indicated that they would override the recommended intervention type if circumstances warranted during an investigation. One example given was if a carrier refused to work with a safety investigator, in which case an offsite investigation would be upgraded to an onsite investigation. Investigators also reported changing the intervention type if additional information uncovered during the investigation merited a change in the intervention level. One investigation manager estimated this occurred in 10–15 percent of investigations. Others said that they don't change the type often, though they did not quantify what "often" meant.

Another respondent indicated that he would change the intervention type based on trends for a carrier. The BASIC scores and therefore whether a BASIC threshold is exceeded can vary from month to month, based on the window of data available. If a specific BASIC for a carrier was close to being deficient, but not over the threshold, the investigator may treat the BASIC as "likely to be deficient" and the type of intervention would be changed accordingly. (This assumes, of course, that the carrier has already been prioritized for intervention based on other BASICs.) On the other hand, a BASIC score may drop below the violation threshold, and the original intervention type would need to be overridden to apply a more appropriate intervention. As one respondent commented, "[t]he data [are] always changing."

Overall, when asked, respondents felt that the CSA 2010 model represented a positive change over the previous SafeStat system. One respondent stated that CSA 2010 was "[t]remendous improvement" over SafeStat. Another wrote that on a 10-point scale, "SafeStat=5 and the CSA SMS = 9+." "The big advantages to CSA is all roadside violations are used. . . . Overall, CSA is superior to SafeStat." Another said that it was "better but could still be improved."

One investigation manager indicated that the new Operational Model was not significantly different from an assignment point of view, pointing to some parallels in the rating system: A/B carriers in SafeStat and "mandatory review" carriers in CSA 2010. And another thought that CSA 2010 was an improvement over SafeStat in identifying problem carriers, but the reason given was that the data window in CSA is shorter than SafeStat. This same respondent expressed concern earlier about the use of older data to identify problem carriers, asserting that violations more than a year old were not reflective of the current status of the carrier. This person said that CSA was better that SafeStat since the data window was 24 rather than 30 months, but still stated that the older data is not useful. In this person's view, only the current state of carrier violations

are useful. Using results from inspections more than a year old "can result in our conducting investigations on carriers with few or no violations in the past year."

On the other hand, drawing on the fact that CSA 2010 significantly increases the range of data from inspections used in evaluating carriers, some respondents pointed to the greater resolution in CSA 2010 in identifying problem areas within a carrier, stating that this allowed investigators to more easily find and work on deficiencies. One characterized CSA 2010 as a "much richer and more defined tool as opposed to SafeStat. . . . [I]t allows a much clearer view of a carrier's compliance breakdowns and enables the Division to task resources accordingly."

Several questions in the survey addressed the process of certain interventions, specifically in off-site investigations where documentation is requested from a carrier and then reviewed for further action. The two major issues addressed in the survey were how well carriers responded to these document requests and what problems arose in processing the documents once they were received.

A wide variety of issues were raised. Document transfers can be bulky, so in some cases the safety investigator will arrange to have the carrier prepare a package with the documents and then have a near-by enforcement officer pick it up. Some respondents simply indicated that the response to document requests was reasonably successful or "[i]t is working fine." Others provided more nuanced responses.

The willingness and ability of carriers to comply with the document requests depends on a number of factors. An offsite request for documents is easier to ignore by the carrier. Several respondents indicated that small carriers do not have someone dedicated to compliance and often they have many other tasks and are not familiar with complying with the request. Timeliness is also an issue for small carriers when they do not have a dedicated compliance person. Companies that have had prior document requests are better able to supply the documents. One investigation manager speculated that "[i]t may be more efficient to take a day or two to do a focused onsite review than to do an offsite review and hope a carrier sends the documents that you need especially in the case of owner/operators." This is an example of changing the intervention type to get a more efficient investigation.

Several emphasized that good compliance on document requests depends on strong and clear communication. In some cases there are problems identifying the exact documents needed because carriers use different terms for the documents. One indicated that generally his Division got good results in document requests. Carriers are "vetted" in advance and those that seem "uncooperative or non-responsive are tasked for onsite focused reviews." This respondent went on to say: "As long as the requests are specific and not too 'broad' document submission [is] highly successful. The use of a document request letter is an effective method to ensure a carrier provides the records and the proper records. Also, establishing a due date with the carrier for receipt of records and including same in the letter [is helpful]."

As might be expected, there was a wide variety of experiences in document requests, and the result really depended on the responsiveness of the carriers. One survey respondent said that he had some good experiences in offsite investigations, including some with small, one or two driver carriers and others with large carriers. But there are also some offsite investigations that

"involved multiple phone calls to the carriers to initially make contact with them, then multiple phone calls to them again when we did not receive any of the requested documents, then multiple phone calls to them when we did not receive the right documents. Then more multiple attempts to contact the carriers in order to close out the reviews." This respondent said the off-site investigation as such did not really save a lot of time, but that his investigators did realize time-saving by "the past practice of almost always completing onsite comprehensive reviews as opposed to onsite focused reviews . . . Now we are seeing [fewer] carriers in CSI that require a full audit (as opposed to focused)."

With respect to the process of requesting documents from the carriers, there was no systematic problem that was identified by the survey respondents. Some experienced good cooperation, while others had specific instances where the process was drawn out by the lack of cooperation from the carrier. Most said that small carriers were less likely to produce the correct documentation in a timely fashion, either because carriers did not understand what was required, were not familiar with the documents, or they were juggling multiple responsibilities. Good, clear, and consistent communication was emphasized, and failing that, converting to an onsite investigation was an option.

In terms of processing the documents once they are received, about half of survey respondents did not report any problems. For the rest, the primary issue raised was the sheer volume of records. Some indicated that sometimes wrong and incomplete files were received. But the larger problem was simply processing the volume of records received. Some indicated problems in working with electronic documents. While printing the documents is expensive, reviewing voluminous electronic documents is time-consuming. Effective review requires organization and arrangement to be efficient. One respondent indicated that using multiple monitors can help in reviewing electronic documents.

Document retention was also reported as a problem with the document requests. The volume can threaten to overwhelm existing IT systems for retention.

One outcome of the investigations is to identify "process breakdowns," which result in noncompliance with motor carrier regulations. The survey respondents were asked to evaluate how well this works and what primary lessons learned.

Overall, respondents felt the process works well and offered many positive comments. One called it a "[v]aluable tool both for enforcement and industry," and said that the safety management wheel graphic "helps [carriers] to visually see how all aspects of a carrier[']s operations have impacts to safety." Another felt that the "Process Breakdown process is tremendous improvement in identifying unsafe motor carrier operations. FMCSA's traditional model focused on identifying violations, particularly 'Critical' and 'Acute' violations. However, this new process has focused the investigation toward identifying any and all performance behaviors which may lead to unsafe operations." The survey respondent said that it was well received by carriers and investigators, one of the best parts of the new CSA 2010 model. "Motor carriers seem more understanding of their actual problems when they are explained not in terms of violations, but in terms of actual causal factors or process breakdowns within their operations."

Another stated that it "needs to be a critical element of almost every intervention. There will be some 'repeat' offenders that do not need a process breakdown, because the carrier has consciously chosen a path of non-compliance. The vast majority of carriers, however, would benefit from this process (which existed before CSA, but was never adequately documented) and it indicates that the Agency's mission is to reduce accidents through whatever means necessary, not always just fines and penalties."

Some respondents said that the process of identifying the critical reasons for noncompliance does not result in changed carrier behavior in all instances, because some carriers are not motivated to comply. "There will be some 'repeat' offenders that do not need a process breakdown, because the carrier has consciously chosen a path of non-compliance." Another respondent indicated that some of the CSI Web site's recommended readings for the carriers on safety and compliance were "lacking." He said that many of the carriers are small ones and the owners would not take the time to read the references, or know what to do with them. "Many of the articles lack any real specific helpful tip in running a motor carrier business." No other respondent made this point, however.

With respect to the interventions that are typically used after an investigation—CSP, NOV, and NOC—survey respondents indicated generally that the range of tools available was appropriate and adequate. No one proposed any other intervention type. And no one argued that any of the interventions was not needed. Most also emphasized that discretion and flexibility in choosing and applying an intervention was very useful.

One respondent said that "experienced safety staff" are "aware of the point at which enforcement should be taken." He also stated that it is appropriate to have alternatives to monetary penalties (by implication the NOV and CSP) when doing nothing is not acceptable. In other words, a monetary penalty through a NOC is not always the most effective way to achieve compliance. A NOV or CSP can also be useful tools for some carriers.

Another commented that "[t]he procedure works, as long as the division office continues to have discretion, as well. Generally, we follow the guidelines, but there are times when we are more inclined to do a NOV while the recommendation is for a NOC. With division discretion, we are able to do that." Several of the respondents emphasized that allowing some discretion and judgment in choosing post-investigations was essential for efficiency and compliance, and that the investigation managers and safety investigators could exercise that judgment. "[T]here are times when we are more inclined to do a NOV while the recommendation is for a NOC." Another respondent agreed, saying that "[i]f the violations do not warrant an NOV or an NOC, and the carrier does not display a positive attitude toward safety, then the CSP is warranted. … It is … good to have many different options at our disposal to get the carrier into compliance." Selecting and employing the most effective tool to bring carriers into compliance, as opposed to punishing them, was the consistent theme in almost all responses.

One objected that for some violations, discretion is taken out of the investigator's hands. For some violations types and for some carrier types (e.g., passenger carriers), current policy prohibits NOVs and requires that they be included in an NOC. More than one respondent said that this inflexibility prevented them from using the NOV to bring about carrier compliance.

"The issue is when current policy (i.e. section 222 or passenger carrier policy that prohibits NOV's) takes the selection from the hands of those involved."

Another respondent raised an almost philosophical question about using an NOC when, in his view, the CSA approach is to identify causal factors and processes for correction.

> *The traditional notice of claim process was aimed at enforcement for "Critical" and "Acute" violations, not process breakdowns. Since the new model is strongly based on identifying causal factors and process breakdowns (which is evident in our usage of all roadside violation data when calculating BASIC measurements, not just out of service data), should we be officially using the same approach when initiating enforcement actions[?] This is particularly concerning since most of the causal factors and process breakdowns discovered during compliance reviews begin or are the result of non-critical and non-acute violations.*

NOVs were generally considered to be effective for most carriers. "The NOV gets their attention and initiates action which is what we want." One respondent observed that NOVs seem particularly effective in the current environment since the carriers have more motivation to avoid the monetary penalty of an NOC. "NOV's require corrective action and have a varied number of uses with or without contact with the subject carrier. They are monitored, tracked and counted which allows the safety investigator to receive credit for work completed. They further demonstrate the Agency's mission and provide the safety investigator the opportunity [to] utilize their onsite skills to determine the correct course of action, or give the carrier a 'chance' to comply prior to assessing a civil penalty."

Others agreed with this observation but said that the NOC also must be included in the range of tools available. "[T]he notice of claim process is traditionally FMCSA's most effective tool for changing behavior." Moreover, an NOV requires resources to track and monitor carriers and can be time consuming. Several said that the NOV requires a commitment of resources to followup to ensure that behavior has changed. One suggested NOVs for violations that are easy to correct. Many violations where policy dictates a NOC could be handled with an NOV. "NOVs force the carrier to correct the violation, while NOCs simply require that they pay money."

Opinion was more varied about the value of CSPs. Some indicated that the value of a CSP was that it forced a carrier to think through his or her operations and to consider what they can do to improve them. "[I]f the carrier is taking the time and effort to put in writing their plan of action as to how they are going to fix the problem or essentially the violation, then there's a presumption that they understand the problem and are more likely to stick with their plan to correct the problem." Moreover, the CSP establishes a basis for further action if the carrier fails to follow through on the CSP. Lessons learned in addressing one area of concern could be used in improving operations in other areas. On the other hand, one respondent pointed out that if the CSP is not mandatory, if "there are no consequences for failing to complete one, the value becomes questionable."

A few of the respondents said that in their experience CSPs were effective. "We have seen a noticeable positive change in some of our carrier's BASIC deficiencies during the 6 to 8 month periods following a compliance review [after] which the carrier received a CSP." More of the

respondents, however, indicated that they did not know whether CSPs were effective, because they had not seen any data on the outcomes. “We have had carriers do [a] CSP but I do not know if this has contributed to the company safety practices or not.” And one respondent reported poor results from CSPs, but primarily because consultants employed by the carriers advise them not to sign the CSP. The CSPs are essentially a voluntary process, and this respondent indicated that the carriers indifferent to compliance are not willing to finalize. In this person’s experience, CSPs are used for carriers with no serious violations, “in other words, the carriers who are substantially in compliance already.”

Overall, the respondents’ views on the CSA 2010 process were generally positive. Most agreed that CSA 2010 represents a significant improvement over the SafeStat process. More carriers are “touched” by CSA than under the previous regime. They are seeing the same carriers as before but also a large number of new carriers, who are discovering that FMCSA is monitoring their safety and is requiring compliance. The respondents generally agree that the system is identifying the right carriers for interventions and provides the right tools to bring carriers into compliance. Several stated, in different contexts, that using all available data on safety is the right thing to do.

Several emphasized the usefulness of preserving flexibility in the local divisions. Their approach is clearly predicated on bringing carriers into compliance and they believe that they have the experience, knowledge, and discretion to choose the best tools to do so. Each of the interventions was regarded as useful, depending on the situation. Some were unwilling to vouch for the effectiveness of certain intervention types, specifically the CSP, but virtually all clearly indicated that each intervention type could be useful in different situations. It also appears from their comments that most regarded most carriers as willing to comply with the regulations, though there are some carriers who are indifferent to compliance and would not comply without coercion.

Several raised questions about the system of assigning priority to carriers for interventions and the relevance of the data used to measure safety. The process of assigning priority order for investigation includes both the time since the last contact with the carrier along with the severity of the violations. Some of the respondents thought that there was too much emphasis on the interval of time since the last carrier contact, and who prefer to put more stress on the current violations of carriers. In that regard, they challenged the relevance of violations committed over a year ago. The investigators would like more safety data and more recent safety data.

Overall, however, the survey of the field staff showed widespread support for the CSA 2010 process. Most would agree that, as one said, “[i]t’s better but could still be improved.”

12. TEST-ONLY STATES

In addition to the four original States in which carriers were randomly allocated into a test group and a control group, five additional States were added as test-only States as the Operational Model Test progressed. As such, all carriers in the newly added States were test carriers and subject to CSA 2010 interventions. According to data recorded in the CSA 2010 files, the five States were added at varying times:

- Minnesota: April 2009.
- Montana: July 2009.
- Delaware, Kansas, Maryland: October 2009.

Since these States were added as the test proceeded, they do not provide 29 months of data, but data from time to inclusion until June 2010. Data collected from the four original States are more consistent with each other since these States were included in the test at the same time, and 29 months of data were recorded for the majority of carriers. However, the same information that is recorded for test carriers in the original four States is also recorded for test carriers in the five test-only States. Therefore, data recorded on test carriers in the new States can in some sense be used to validate certain findings presented for carriers in the original four States.

12.1 SAFETY MEASUREMENT THRESHOLDS EXCEEDED IN TEST-ONLY STATES

Inspection of SMS data on the test-only States suggests that by January 2010, all test-only carriers had been added to the database. After that time, until June, 2010, the total number of test-only carriers remains fairly stable. Table 60 shows the distribution of test-only carriers by number of SMS thresholds exceeded during the month of January 2010. Due to some differences, Minnesota and Montana have been categorized into one group, and Delaware, Kansas, and Maryland into another. For Minnesota and Montana, results are very similar to results presented in Table 7 from the original four States. The majority of carriers did not exceed any SMS thresholds, while 6.0 percent exceeded at least one. For the other three States, the results are a little different. In Delaware, Kansas, and Maryland combined, about 4.1 percent of carriers exceeded at least one threshold.

Table 60. Number of SMS Thresholds Exceeded (Test-Only States, January 2010)

Thresholds Exceeded	Carriers (MN, MT)	Percent	Carriers (DE, KS, MD)	Percent
0	22,230	94.0%	37,590	95.9%
1	1,027	4.3%	1,176	3.0%
2	268	1.1%	325	0.8%
3	83	0.4%	86	0.2%
4	31	0.1%	28	0.1%
5+	7	<0.1%	10	<0.1%
Total	**23,646**	**100.0%**	**39,215**	**100.0%**

When carriers exceed any threshold, they usually exceed only one. Table 61 shows the percentage of carriers that exceeded each of the BASIC thresholds for carriers that exceeded one threshold. Results are also broken out by two groups of States. As in the original four States, Fatigued Driving and Vehicle Maintenance have the largest percentages. However, in Minnesota and Montana, the percentage for Fatigued Driving is 43.8 percent (compared to 35.2 percent for all carriers in the original four test states). In Delaware, Kansas, and Maryland, the largest percentage of carriers exceeding a threshold is 28.5 percent for the Vehicle Maintenance BASIC. The Driver Fitness BASIC also shows some differences, with 1.7 percent of carriers exceeding that threshold in Minnesota and Montana, but 13.8 percent in the other three States.

Table 61. SMS Threshold Exceeded for Carriers Exceeding One Threshold (Test-Only States, January 2010)

Threshold Exceeded	(MN, MT) Carriers	(MN, MT) Percent	(DE, KS, MD) Carriers	(DE, KS, MD) Percent
1–Unsafe Driving	94	9.2%	124	10.5%
2–Fatigued Driving	450	43.8%	285	24.2%
3–Driver Fitness	17	1.7%	162	13.8%
4–Controlled Substance/ Alcohol	4	0.4%	21	1.8%
5–Vehicle Maintenance	211	20.5%	335	28.5%
6–Improper Loading/Cargo Securement	201	19.6%	188	16.0%
7–Crash Indicator	50	4.9%	61	5.2%
Total	**1,027**	**100.0%**	**1,176**	**100.0%**

12.2 INTERVENTIONS THAT CARRIERS RECEIVED IN TEST-ONLY STATES

Table 62 shows the percentage of carriers with interventions by number of interventions received for the test-only carriers in the five States. There were 4,763 carriers that received 5,957 interventions at any time after these States were added to the test. Among carriers that received interventions, 78.8 percent received one. An additional 17.8 percent received two interventions. Relatively few carriers received more than two interventions, with approximately 3.4 percent receiving three or more.

Table 62. Percent of Carriers Receiving Interventions by Number of Interventions, Test-Only States

Number of Interventions	Carriers	Total Interventions	Percent Carriers
1	3,755	3,755	78.8%
2	847	1,694	17.8%
3	139	417	2.9%
4	19	76	0.4%
5	3	15	0.1%
Total	**4,763**	**5,957**	**100.0%**

Table 63 shows the percentage of carriers that received interventions when the last intervention is closed/completed (all tasks completed and the intervention closed). In this case, results are restricted to the 3,703 carriers that received 4,459 interventions.

Table 63. Percent of Carriers Receiving Interventions by Number of Interventions, Last Intervention Closed/Completed, Test-Only States

Number of Interventions	Carriers	Total Interventions	Percent Carriers
1	3,069	3,069	82.9%
2	527	1,054	14.2%
3	95	285	2.6%
4	9	36	0.2%
5	3	15	0.1%
Total	**3,703**	**4,459**	**100.0%**

Table 64 shows the distribution of carriers with interventions by the last closed/completed intervention. In some sense, these carriers represent a group for potential evaluation since the last one is closed/completed. However, closed/completed only means that the safety investigator has completed tasks associated with a particular intervention, and not that any safety issues have been necessarily resolved. A carrier is still likely being monitored for potential additional interventions.

Among carriers with one closed/completed intervention, 76.3 percent received warning letters. An additional 9.6 percent received the onsite focused investigation, 8.0 percent received the onsite comprehensive, and 5.5 percent received the offsite investigation. For carriers with two or more interventions in which the last intervention is closed/completed, the NOC is most common. Among carriers with two interventions, the NOC represents 26.9 percent, and the onsite focused 25.4 percent. Percentages associated with the CSP and the NOV are relatively small. This may partly be due to the test-only States being added at later times during the Operational Model Test, and the CSP and NOV tend to be terminal interventions.

Table 64. Distribution of Carriers With Interventions by Last Closed/Completed Intervention and Number of Interventions, Test-Only States

Last Intervention Closed/Completed	1 Intervention Carriers (Percent)	2 Interventions Carriers (Percent)	3 Interventions Carriers (Percent)	>3 Interventions Carriers (Percent)
Warning Letter	2,341 (76.3)	108 (20.5)	4 (4.2)	1 (8.3)
CSA	0 (90.0)	30 (5.7)	8 (8.4)	1 (8.3)
NOV	4 (0.1)	14 (2.7)	3 (3.2)	0 (0.0)
NOC	7 (0.2)	142 (26.9)	57 (60.0)	9 (75.0)
Offsite Investigation	169 (5.5)	44 (8.3)	4 (4.2)	0 (0.0)
Onsite Focused Investigation	296 (9.6)	134 (25.4)	14 (14.7)	1 (8.3)
Onsite Comprehensive Investigation	246 (8.0)	52 (9.9)	5 (5.3)	0 (0.0)
Followup Verification	6 (0.2)	3 (0.6)	0 (0.0)	0 (0.0)
Total	**3,069 (100.0)**	**527 (100.0)**	**95 (100.0)**	**12 (100.0)**

12.3 CARRIERS TOUCHED BY CSA 2010 IN TEST-ONLY STATES

Table 65 shows the annual percentage of carriers with interventions based on the six months from January, 2010, through June, 2010. This 6-month period was chosen because all five test-only States were participating fully in the test during this time, according to information recorded in the CSA 2010 database. The total number of carriers with interventions is 1,978 and the average number over 6 months is 47,099. The annual percentage expected to be touched is 8.4 percent. This compares to 9.9 percent shown in Table 14 for carriers in the original four States.

Table 65. Annual Percentage of Test Group Carriers With Recent Activity Touched by Interventions (Test-Only States DE, KS, MD, MN, MT—6 Months)

Average Number of Test Carriers	Total Interventions	Total Carriers With Interventions	Annual Number of Carriers With Interventions	Annual Percent of Carriers With Interventions
47,099	2,322	1,978	3,956	8.4%

13. SUMMARY AND DISCUSSION

In accordance with its primary mission to reduce crashes, injuries, and fatalities involving large trucks and buses, the FMCSA initiated the CSA 2010 Operational Model Test. Major goals of the test were to improve the current process used to monitor and assess the safety performance of motor carriers and drivers operating on the Nation's highways. The CSA 2010 program focused on initiating contact with more carriers and drivers, development of a new measurement system to replace SafeStat, application of a wider range of progressive interventions to correct high-risk behavior, and more efficient use of Agency resources.

Originally, the test was conducted in Colorado, Georgia, Missouri, and New Jersey. Carriers in these States were randomly divided into a test group that was subject to the provisions of the new CSA Operational Model, and a control group that would continue to be monitored by the Agency's current process. For the four original States, the test ran for 29 months from February, 2008, through June, 2010. However, five additional States (Montana, Minnesota, Maryland, Kansas, and Delaware) were phased into the program as *test-only* States at various times as the test proceeded.

The CSA process begins for a carrier when a carrier exceeds one of the BASIC thresholds of the new SMS. The carrier is then subject to any number of various interventions, depending on the type and number of thresholds exceeded. The intervention process is designed to address the specific behavior that precipitated the intervention, namely, exceeding at least one SMS threshold, so that further interventions are not necessary.

This report is an evaluation of the CSA 2010 Operational Model Test. The evaluation focuses on the key components of the Operational Model. In particular, the thresholds exceeded in terms of type and frequency, the interventions received in terms of type and frequency, and the number and percentage of carriers *touched* under the new model are calculated and reported. Intervention cycles and patterns were explored and effectiveness of interventions was assessed by comparing test carriers that received CSA 2010 interventions to control carriers that did not receive CSA interventions, and to control carriers that received traditional CRs. Costs to the agency of performing CSA2010 interventions and total amounts claimed were investigated and compared to the current process of conducting CRs. The new SMS that is used to rank a carrier's safety performance in the seven BASICs was evaluated largely by assessing associations between percentile scores and crash rates. Comparisons were made with the current model under SafeStat, and quality of the MCMIS data files used by both measurement systems was assessed. Finally, results from a survey completed by field staff participating in the Operational Model Test were presented to determine which aspects of CSA 2010 worked well and which did not.

Overall, approximately 6.1 percent of carriers exceeded at least one of the SMS thresholds. Approximately 4.5 percent exceeded one threshold and 1.3 percent exceeded two. Only about 0.3 percent exceeded more than two. Vehicle Maintenance and Fatigued Driving thresholds were most likely exceeded. For carriers that exceeded only one threshold, these two BASICs accounted for about 73 percent of the total. When multiple thresholds were exceeded, in addition to these two, the Unsafe Driving BASIC tended to be included.

For carriers receiving interventions, about 50 percent received one. Among intervention patterns, the Warning Letter, with no interventions following thereafter, was by far the most likely. About one-third of all intervention patterns consisted of a warning letter followed by no other interventions (Figure 24). Most patterns were characterized by one or two interventions. Often they involved one of the investigations (offsite, onsite focused, onsite comprehensive) either as the first intervention, or following the warning letter. The CSP and the NOC tend to be terminal interventions, and usually occurred as the second or third intervention. The NOV was rarely used.

Annually, it is estimated that CSA interventions will touch approximately 6.3 percent of the carrier population (based on all interventions, including warning letters). This compares to about 2.2 percent of carriers that receive full CRs under the current process. Therefore, the number of carriers *touched* by CSA on an annual basis is approximately 6.3/2.2 = 2.9 times greater than the current system based on CRs alone. When restricted to carriers with recent activity, CSA is expected to touch about 9.9 percent of the population. (Excluding warning letters, the percentage of carriers expected to be investigated annually under the CSA program is about 6.2 percent.) The corresponding percent of "recent activity" carriers with CRs is 3.2 percent. The ratio 9.9/3.2 = 3.1 remains fairly constant.

The effectiveness of CSA 2010 interventions was assessed by comparing test carriers that received CSA 2010 interventions to control carriers that did not. Three control groups were considered. One control group consisted of control carriers without CRs. In terms of safety risk, this group was considered to be a mild group. Another group consisted of carriers with CRs. This group was considered to be representative of high-risk carriers. The third group was sampled from control carriers to have safety characteristics similar to test carriers prior to evaluation. Carriers were then followed over a 12-month period and evaluated on the percentage of SMS thresholds exceeded. The procedure is described in detail in section 8 and Figure 15 depicts a timeline of events.

It appears that the FMCSA was most successful in bringing carriers under SMS thresholds in cases where a Warning Letter was followed by no other interventions. After 12 months of followup, approximately 17 percent of test carriers were still exceeding at least one SMS threshold, compared to about 45 percent of control carriers matched to test carriers. Carriers addressed with a "warning letter only" tended to have milder violations, so that the warning letter sufficed to resolve the problem. As discussed above, this intervention pattern was also found to be the most prevalent, accounting for one-third of the patterns and giving the largest sample size of carriers for evaluation.

The effectiveness of the investigations (offsite, onsite focused, onsite comprehensive) was also examined for carriers with one intervention. A similar pattern emerged. There appeared to be a lag time of about eight months before carriers with these interventions improved. Once they did show signs of improvement, however, generally a smaller percentage of them exceeded the BASIC threshold values than their counterparts in the control groups. For example, for the onsite focused investigation, after 12 months of followup, 40.6 percent of test carriers were still exceeding at least one SMS threshold, compared to about 56.7 percent of control carriers matched to test carriers, and about 60 percent for control group carriers receiving traditional

CRs. Unlike the "warning letter only" pattern, however, sample sizes for carriers with interventions based on investigations are smaller.

For carriers with multiple interventions, it is more difficult to assess the effects that interventions had on safety behavior. For example, intervention patterns that ended in an NOC generally took about 1 year to be closed/completed (all tasks completed and the intervention closed). In addition, these carriers tend to be high-risk carriers from the beginning and during the investigation process. Therefore, many of these carriers still exceeded SMS thresholds after the investigation became closed/completed, and the test group does not show lower percentages of thresholds exceeded than a control group even after matching controls on crash rates and BASIC scores. However, for carriers with interventions ending in a NOC, Figure 25 shows test carriers initially having high percentages of any thresholds exceeded and improving relative to the control group with CRs and matched to the test group after 12 months of followup.

Costs to FMCSA of performing CSA 2010 interventions on carriers in the Operational Model Test were investigated. All CSA2010 interventions considered were those that were classified as closed/completed to ensure that all tasks performed during a particular intervention have been completed. For comparison, costs to the Agency of performing CRs on control group carriers under the current enforcement program were also calculated. The estimated average cost per carrier for all interventions performed is $754, while the median cost is $590. When restricting to SafeStat A/B carriers, the average cost is $1,032 and the median cost is $824. For comparison, the average cost for non-test carriers receiving CRs is $1,438 and the median is $1,058.

In addition to the costs to FMCSA of performing interventions on carriers, dollar amounts claimed from test and control group carriers were also calculated. During the 29 month evaluation period of the Operational Model Test, 720 claims were made on test carriers and 640 claims were made on control carriers in the original four states. The average amount per claim was $5,016 for test carriers and $6,296 for control carriers. Because a few carriers had very large claims, the average does not reflect the typical claim, and the median is a better statistic for comparison. For test carriers the median is $2,200 and for control carriers the median is $2,480. The conclusion is that there is not a significant difference in the experience between test and control carriers with respect to claims made.

The SMS was evaluated to determine to what extent the SMS identifies unsafe carriers. Crash rates were calculated for carriers that exceeded the BASIC thresholds, and were compared to the crash rates of carriers that did not exceed any BASIC thresholds. Crash rates were also calculated for carriers identified under the current SafeStat system and compared to those under the SMS. In addition, scatter plots were made to assess associations between BASIC percentiles and crash rates. To provide a large sample of carriers, crash rates were calculated using 473,847 carriers not participating in the CSA 2010 test.

For all BASICs, crash rates were higher for carriers exceeding SMS thresholds than for carriers not exceeding thresholds. The crash rate was highest for carriers exceeding the Unsafe Driving threshold. Rates were also high for the Fatigued Driving BASIC and the Controlled Substance and Alcohol BASIC. The SMS also identified many more carriers for intervention than did SafeStat. Scatter plots indicate that all of the BASIC measures have positive associations with

crash rates, except for the Driver Fitness and Loading/Cargo Securement BASICs. Excluding the Crash Indicator, the Unsafe Driving BASIC has the strongest association with crash rates.

A survey of field staff participating in the Operational Model Test was conducted to determine which aspects of CSA 2010 worked well and which did not. The purpose was to collect their insights and perspective on both the process and results of the new system of identifying unsafe carriers and intervening. Responses were received from eight States, including all four of the original test States: New Jersey, Colorado, Georgia, and Missouri. Overall, the evaluation of CSA 2010 was quite positive, in terms of identifying the right carriers for interventions and having the right tools to improve carrier behavior. There were some critiques of the method of prioritizing carriers for interventions and questions about the quality of the data used to score the BASICs (mostly having to do with the timeliness of the data) but, overall the people who were actually operating the new system believe that the new CSA model represents a significant improvement over the prior SafeStat system.

APPENDIX A: SUPPORTING DATA FOR PLOTS WITH 12 MONTHS FOLLOWUP (SECTION 7)

Table 66. BASIC 1: Unsafe Driving Test

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	132	132	100.0%
1	116	132	87.9%
2	93	128	72.7%
3	77	121	63.6%
4	60	111	54.1%
5	47	106	44.3%
6	36	98	36.7%
7	25	91	27.5%
8	23	84	27.4%
9	16	81	19.8%
10	12	78	15.4%
11	12	74	16.2%
12	10	69	14.5%

Table 67. BASIC 1: Unsafe Driving Test
Control Without CR

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	667	667	100.0%
1	570	667	85.5%
2	524	651	80.5%
3	458	634	72.2%
4	406	611	66.4%
5	363	591	61.4%
6	304	573	53.1%
7	273	554	49.3%
8	245	537	45.6%
9	227	517	43.9%
10	210	488	43.0%
11	196	463	42.3%
12	168	449	37.4%

Table 68. BASIC 2: Fatigued Driving Test

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	693	693	100.0%
1	588	693	84.8%
2	517	677	76.4%
3	444	650	68.3%
4	391	605	64.6%
5	337	588	57.3%
6	289	573	50.4%
7	255	551	46.3%
8	224	525	42.7%
9	200	504	39.7%
10	181	486	37.2%
11	146	465	31.4%
12	110	447	24.6%

Table 69. BASIC 2: Fatigued Driving Control Without CR

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	2037	2037	100.0%
1	1733	2037	85.1%
2	1575	1991	79.1%
3	1423	1942	73.3%
4	1287	1885	68.3%
5	1210	1845	65.6%
6	1102	1801	61.2%
7	983	1758	55.9%
8	915	1714	53.4%
9	826	1661	49.7%
10	790	1615	48.9%
11	719	1560	46.1%
12	616	1516	40.6%

Table 70. BASIC 3: Driver Fitness Test

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	121	121	100.0%
1	107	121	88.4%
2	89	118	75.4%
3	73	109	67.0%
4	56	103	54.4%
5	51	100	51.0%
6	40	96	41.7%
7	33	94	35.1%
8	26	91	28.6%
9	25	86	29.1%
10	21	82	25.6%
11	21	75	28.0%
12	14	73	19.25%

Table 71. BASIC 3: Driver Fitness Control Without CR

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	454	454	100.0%
1	384	454	84.6%
2	329	440	74.8%
3	308	428	72.0%
4	282	420	67.1%
5	250	410	61.0%
6	222	399	55.6%
7	206	393	52.4%
8	189	381	49.6%
9	169	373	45.3%
10	154	365	42.2%
11	143	359	39.8%
12	123	344	35.8%

Table 72. BASIC 5: Vehicle Maintenance Test

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	692	692	100.0%
1	583	692	84.2%
2	514	673	76.4%
3	451	656	68.8%
4	373	620	60.2%
5	329	598	55.0%
6	286	578	49.5%
7	245	556	44.1%
8	209	531	39.4%
9	176	508	34.6%
10	154	484	31.8%
11	128	468	27.4%
12	111	442	25.1%

Table 73. BASIC 5: Vehicle Maintenance Control Without CR

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	2216	2216	100.0%
1	1894	2216	85.5%
2	1716	2175	78.9%
3	1552	2113	73.5%
4	1401	2049	68.4%
5	1322	2010	65.8%
6	1240	1968	63.0%
7	1139	1917	59.4%
8	1060	1875	56.5%
9	991	1832	54.1%
10	894	1767	50.6%
11	838	1706	49.1%
12	779	1658	47.0%

Table 74. BASIC 6: Improper Loading/Cargo Securement Test

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	188	188	100.0%
1	162	188	86.2%
2	145	184	78.8%
3	112	178	62.9%
4	100	169	59.2%
5	87	162	53.7%
6	78	157	49.7%
7	76	155	49.0%
8	64	150	42.7%
9	65	147	44.2%
10	60	142	42.3%
11	55	138	39.9%
12	48	135	35.6%

Table 75. BASIC 6: Improper Loading/Cargo Securement Control Without CR

Month	BASIC Threshold Exceeded	Total Carriers	Percent
0	685	685	100.0%
1	584	685	85.3%
2	527	673	78.3%
3	480	662	72.5%
4	446	649	68.7%
5	411	637	64.5%
6	386	629	61.4%
7	362	619	58.5%
8	342	609	56.2%
9	331	600	55.2%
10	320	588	54.4%
11	309	577	53.6%
12	282	565	49.9%

[This page intentionally left blank.]

APPENDIX B: SUPPORTING DATA FOR PLOTS WITH 12 MONTHS FOLLOWUP (SECTION 8)

Table 76. Intervention Pattern: Warning Letter Only Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	1,749	1,749	100.0%
1	1,261	1,749	72.1%
2	1,064	1,683	63.2%
3	891	1,583	56.3%
4	774	1,525	50.8%
5	652	1,428	45.7%
6	590	1,423	41.5%
7	510	1,363	37.4%
8	422	1,276	33.1%
9	371	1,227	30.2%
10	292	1,114	26.2%
11	219	1,051	20.5%
12	181	1,048	17.35%

Table 77. Intervention Pattern: Warning Letter Only Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	4,281	4,281	100.0%
1	3,508	4,281	81.9%
2	3,211	4,189	76.7%
3	2,945	4,081	72.2%
4	2,747	4,007	68.6%
5	2,572	3,921	65.6%
6	2,349	3,840	61.2%
7	2,216	3,758	59.0%
8	2,086	3,677	56.7%
9	1,945	3,586	54.2%
10	1,824	3,481	52.4%
11	1,675	3,394	49.4%
12	1,550	3,295	47.0%

Table 78. Intervention Pattern: Warning Letter Only
Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 79. Intervention Pattern: Warning Letter Only
Control Matched to Test

Month	Any BASIC Threshold Exceeded	Test Total Carriers	Percent
0	1,000	1,000	100.0%
1	805	1,000	80.5%
2	739	963	76.7%
3	675	931	72.5%
4	623	905	68.8%
5	592	882	67.1%
6	537	860	62.4%
7	498	829	60.1%
8	457	805	56.8%
9	425	779	54.6%
10	394	755	52.2%
11	361	731	49.4%
12	316	699	45.2%

Table 80. Intervention Pattern: Offsite Investigation—One Intervention Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	68	68	100.0%
1	51	68	75.0%
2	48	67	71.6%
3	49	67	73.1%
4	47	65	72.3%
5	43	63	68.3%
6	42	62	67.7%
7	40	59	67.8%
8	37	57	64.9%
9	30	54	55.6%
10	30	53	56.6%
11	29	53	54.7%
12	22	49	44.9%

Table 81. Intervention Pattern: Offsite Investigation—One Intervention Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	4,008	4,008	100.0%
1	2,750	4,008	68.6%
2	2,575	3,921	65.7%
3	2,352	3,840	61.3%
4	2,218	3,757	59.0%
5	2,087	3,676	56.8%
6	1,948	3,585	54.3%
7	1,828	3,480	52.5%
8	1,680	3,392	49.5%
9	1,555	3,293	47.2%
10	1,472	3,179	46.3%
11	1,381	3,101	44.5%
12	1,293	3,003	43.1%

Table 82. Intervention Pattern: Offsite Investigation—One Intervention Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 83. Intervention Pattern: Offsite Investigation—One Intervention Control Matched to Test

Test Month	Any BASIC Threshold Exceeded	Total Carriers	Test Percent
0	1,241	1,241	100.0%
1	960	1,241	77.4%
2	890	1,222	72.8%
3	843	1,198	70.4%
4	777	1,168	66.5%
5	732	1,151	63.6%
6	684	1,124	60.9%
7	651	1,106	58.9%
8	627	1,075	58.3%
9	595	1,053	56.5%
10	566	1,027	55.1%
11	547	987	55.4%
12	515	953	54.0%

Table 84. Intervention Pattern: Onsite Focused Investigation—One Intervention Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	84	84	100.0%
1	72	84	85.7%
2	69	83	83.1%
3	63	81	77.8%
4	62	79	78.5%
5	56	73	76.7%
6	51	67	76.1%
7	42	58	72.4%
8	38	55	69.1%
9	30	49	61.2%
10	25	45	55.6%
11	22	41	53.7%
12	13	32	40.6%

Table 85. Intervention Pattern: Onsite Focused Investigation—One Intervention Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	4,008	4,008	100.0%
1	2,750	4,008	68.6%
2	2,575	3,921	65.7%
3	2,352	3,840	61.3%
4	2,218	3,757	59.0%
5	2,087	3,676	56.8%
6	1,948	3,585	54.3%
7	1,828	3,480	52.5%
8	1,680	3,392	49.5%
9	1,555	3,293	47.2%
10	1,472	3,179	46.3%
11	1,381	3,101	44.5%
12	1,293	3,003	43.1%

Table 86. Intervention Pattern: Onsite Focused Investigation—One Intervention Control With CR

Month	Any BASIC Threshold Exceeded	CR Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 87. Intervention Pattern: Onsite Focused Investigation—One Intervention Control Matched to Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	837	837	100.0%
1	635	837	75.9%
2	615	822	74.8%
3	557	799	69.7%
4	525	779	67.4%
5	497	759	65.5%
6	480	738	65.0%
7	456	720	63.3%
8	437	693	63.1%
9	417	670	62.2%
10	392	651	60.2%
11	359	620	57.9%
12	337	594	56.7%

Table 88. Intervention Pattern: Onsite Comprehensive Investigation—One Intervention Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	33	33	100.0%
1	27	33	81.8%
2	24	30	80.0%
3	22	29	75.9%
4	17	26	65.4%
5	16	24	66.7%
6	14	23	60.9%
7	16	23	69.6%
8	13	20	65.0%
9	12	20	60.0%
10	9	18	50.0%
11	8	16	50.0%
12	5	11	45.5%

Table 89. Intervention Pattern: Onsite Comprehensive Investigation—One Intervention Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	3,772	3,772	100.0%
1	2,228	3,772	59.1%
2	2,093	3,691	56.7%
3	1,952	3,600	54.2%
4	1,833	3,493	52.5%
5	1,685	3,405	49.5%
6	1,560	3,305	47.2%
7	1,474	3,189	46.2%
8	1,383	3,110	44.5%
9	1,294	3,011	43.0%
10	1,219	2,925	41.7%
11	1,145	2,825	40.5%
12	1,085	2,716	39.9%

Table 90. Intervention Pattern: Onsite Comprehensive Investigation—One Intervention Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 91. Intervention Pattern: Onsite Comprehensive Investigation—One Intervention Test Control Matched to Test

Month	Any BASIC Threshold Exceeded	Test Total Carriers	Test Percent
0	193	193	100.0%
1	153	193	79.3%
2	148	187	79.1%
3	136	181	75.1%
4	125	173	72.3%
5	123	166	74.1%
6	119	163	73.0%
7	112	161	69.6%
8	104	155	67.1%
9	98	144	68.1%
10	96	138	69.6%
11	90	132	68.2%
12	86	125	68.8%

Table 92. Intervention Pattern: Warning Letter—Investigation Test (Two Interventions)

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	376	376	100.0%
1	279	376	74.2%
2	252	350	72.0%
3	214	324	66.0%
4	182	290	62.8%
5	171	266	64.3%
6	143	241	59.3%
7	120	215	55.8%
8	101	201	50.2%
9	80	180	44.4%
10	76	162	46.9%
11	63	142	44.4%
12	54	130	41.5%

Table 93. Intervention Pattern: Warning Letter—Investigation Test (Two Interventions) Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	3,146	3,146	100.0%
1	1,396	3,146	44.4%
2	1,309	3,047	43.0%
3	1,234	2,961	41.7%
4	1,161	2,861	40.6%
5	1,099	2,751	39.9%
6	945	2,461	38.4%
7	888	2,401	37.0%
8	843	2,313	36.4%
9	790	2,240	35.3%
10	720	2,139	33.7%
11	657	2,031	32.3%
12	603	1,891	31.9%

Table 94. Intervention Pattern: Warning Letter—Investigation Test (Two Interventions)
Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 95. Intervention Pattern: Warning Letter—Investigation Test (Two Interventions)
Control Matched to Test

Test Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	796	796	100.0%
1	479	796	60.2%
2	444	773	57.4%
3	424	744	57.0%
4	377	640	58.9%
5	357	616	58.0%
6	321	574	55.9%
7	291	543	53.6%
8	257	499	51.5%
9	236	470	50.2%
10	200	425	47.1%
11	179	392	45.7%
12	162	351	46.2%

Table 96. Intervention Pattern: Warning Letter—Investigation—CSP Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	385	385	100.0%
1	299	385	77.7%
2	288	370	77.8%
3	244	342	71.3%
4	218	314	69.4%
5	207	299	69.2%
6	188	285	66.0%
7	166	264	62.9%
8	147	242	60.7%
9	137	228	60.1%
10	120	210	57.1%
11	106	194	54.6%
12	92	185	49.7%

Table 97. Intervention Pattern: Warning Letter—Investigation—CSP Test Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	3,432	3,432	100.0%
1	1,700	3,432	49.5%
2	1,572	3,331	47.2%
3	1,486	3,213	46.2%
4	1,392	3,134	44.4%
5	1,304	3,035	43.0%
6	1,229	2,949	41.7%
7	1,155	2,849	40.5%
8	1,093	2,739	39.9%
9	941	2,450	38.4%
10	886	2,391	37.1%
11	842	2,303	36.6%
12	789	2,231	35.4%

Table 98. Intervention Pattern: Warning Letter—Investigation—CSP Test Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 99. Intervention Pattern: Warning Letter—Investigation—CSP Test Control Matched to Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	986	986	100.0%
1	667	986	67.6%
2	622	956	65.1%
3	583	923	63.2%
4	532	847	62.8%
5	504	815	61.8%
6	464	777	59.7%
7	429	744	57.7%
8	401	704	57.0%
9	358	656	54.6%
10	331	613	54.0%
11	300	573	52.4%
12	261	512	51.0%

Table 100. Intervention Pattern: Interventions Terminating in NOC Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	267	267	100.0%
1	232	267	86.9%
2	204	247	82.6%
3	187	229	81.7%
4	168	209	80.4%
5	143	187	76.5%
6	126	168	75.0%
7	109	159	68.6%
8	96	150	64.0%
9	87	134	64.9%
10	79	120	65.8%
11	66	106	62.3%
12	55	95	57.9%

Table 101. Intervention Pattern: Interventions Terminating in NOC Test Control Without CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	3,432	3,432	100.0%
1	1,700	3,432	49.5%
2	1,572	3,331	47.2%
3	1,486	3,213	46.2%
4	1,392	3,134	44.4%
5	1,304	3,035	43.0%
6	1,229	2,949	41.7%
7	1,155	2,849	40.5%
8	1,093	2,739	39.9%
9	941	2,450	38.4%
10	886	2,391	37.1%
11	842	2,303	36.6%
12	789	2,231	35.4%

Table 102. Intervention Pattern: Interventions Terminating in NOC Test Control With CR

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	800	800	100.0%
1	613	800	76.6%
2	558	750	74.4%
3	513	698	73.5%
4	455	644	70.7%
5	414	594	69.7%
6	360	538	66.9%
7	341	505	67.5%
8	301	468	64.3%
9	268	421	63.7%
10	227	368	61.7%
11	198	320	61.9%
12	165	278	59.4%

Table 103. Intervention Pattern: Interventions Terminating in NOC Test Control Matched Test

Month	Any BASIC Threshold Exceeded	Total Carriers	Percent
0	420	420	100.0%
1	303	420	72.1%
2	283	408	69.4%
3	265	393	67.4%
4	221	331	66.8%
5	213	316	67.4%
6	204	305	66.9%
7	194	289	67.1%
8	174	277	62.8%
9	160	259	61.8%
10	151	246	61.4%
11	140	230	60.9%
12	133	219	60.7%

APPENDIX C: BOX PLOTS OF MATCHED DISTRIBUTIONS AND CRASH RATES (SECTION 8)

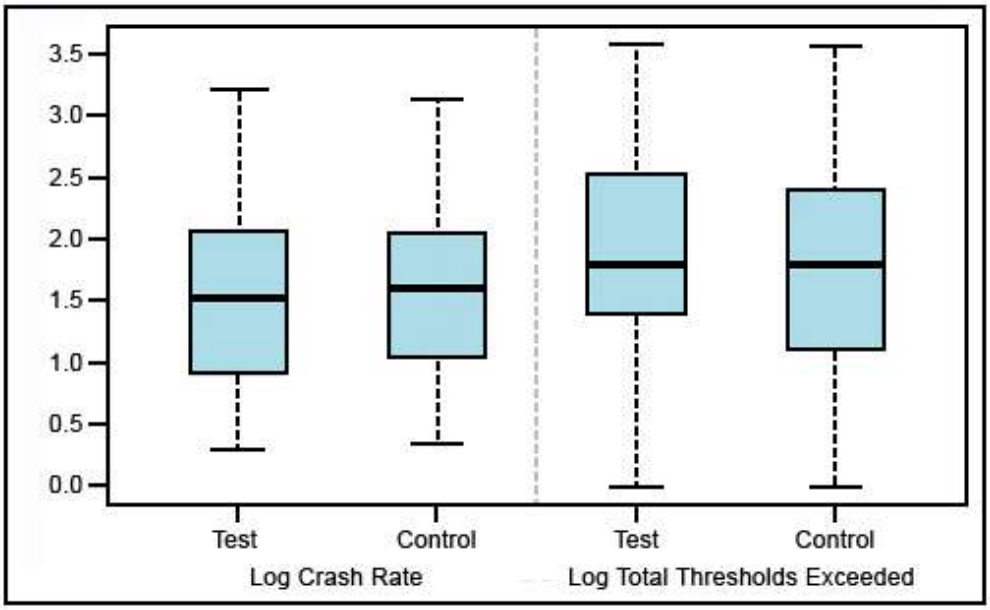

Figure 27. Onsite Focused Investigation

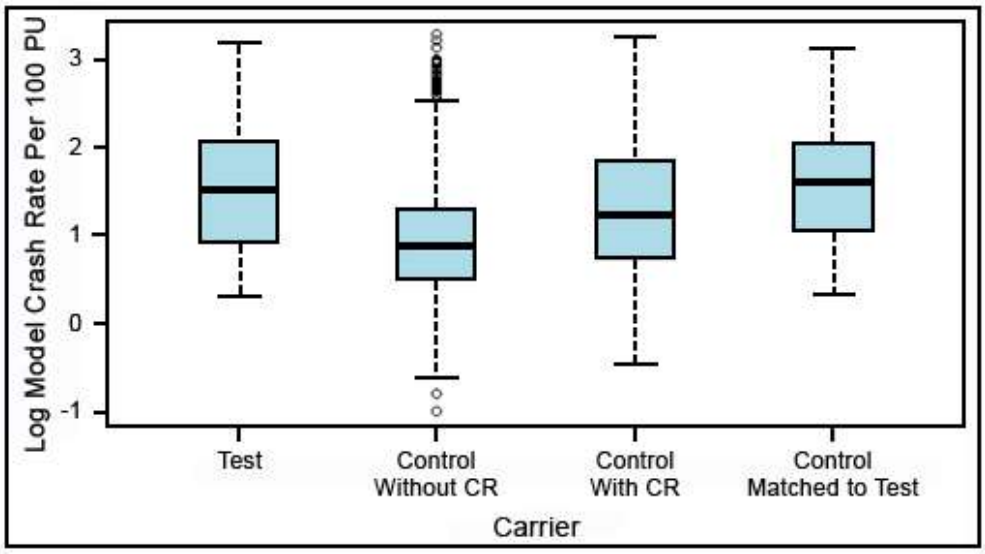

Figure 28. Onsite Focused Investigation

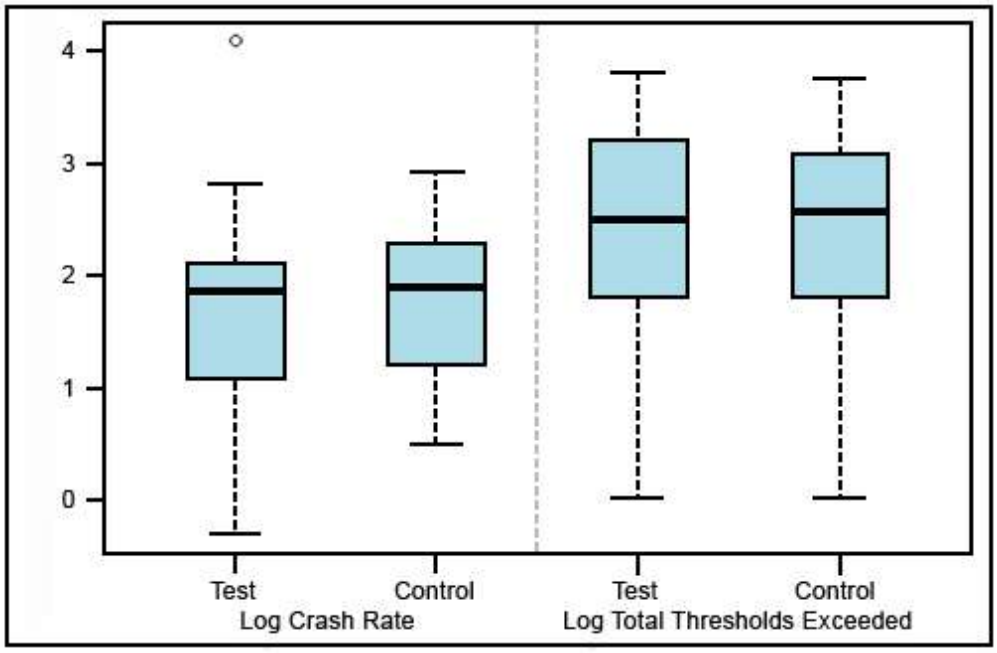

Figure 29. Onsite Comprehensive Investigation

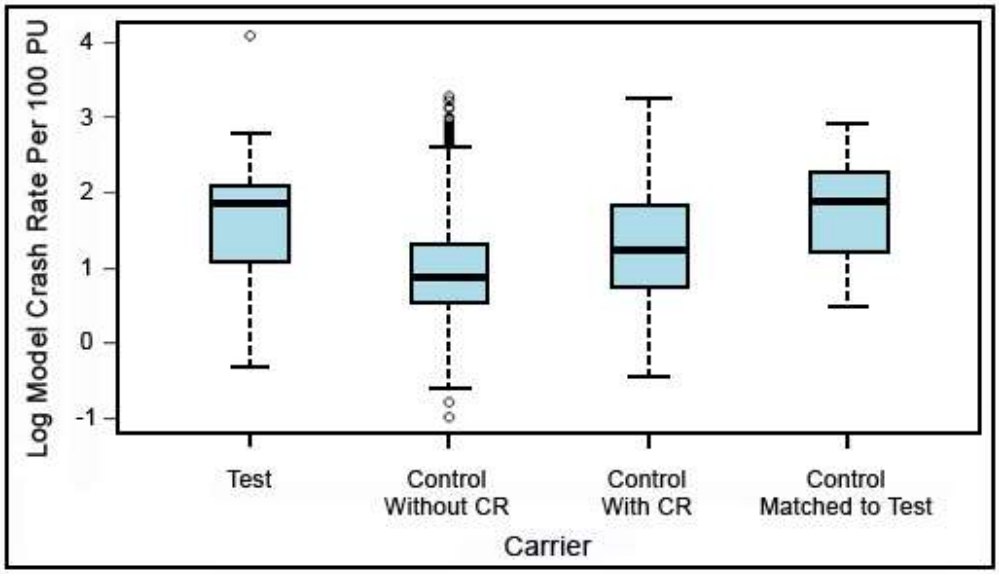

Figure 30. Onsite Comprehensive Investigation

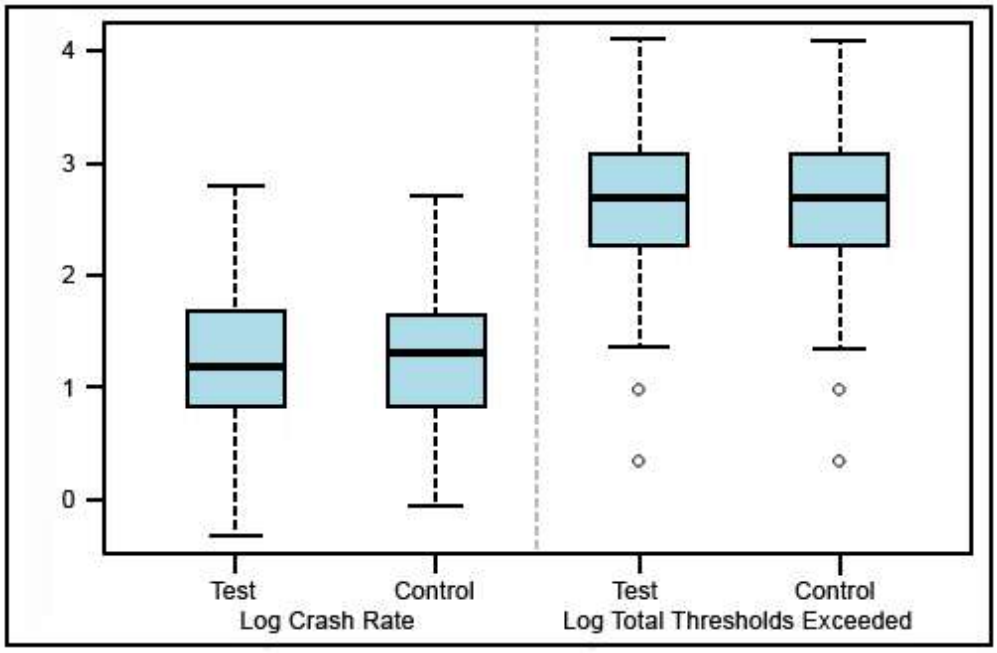

Figure 31. Warning Letter—Investigation

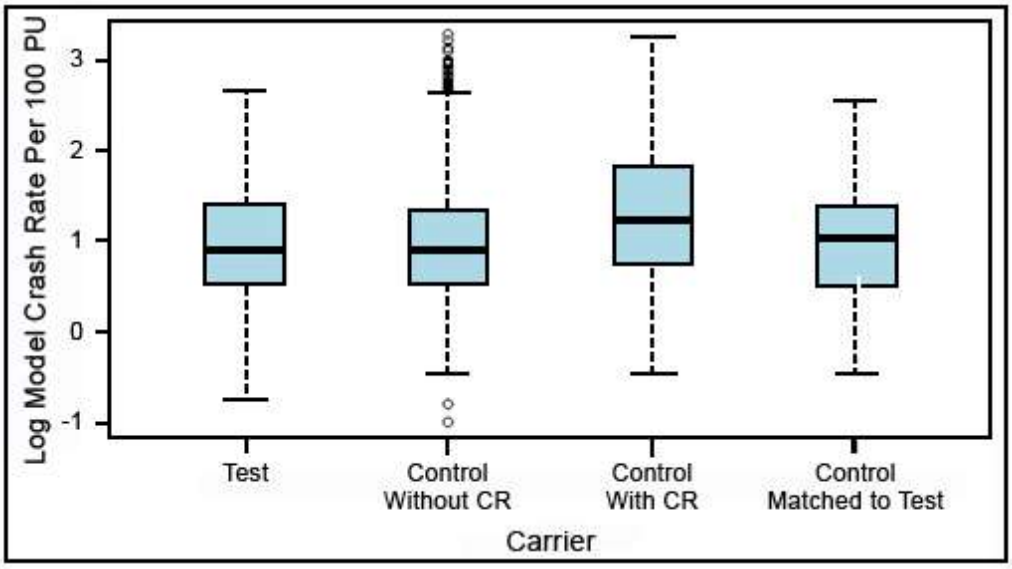

Figure 32. Warning Letter—Investigation

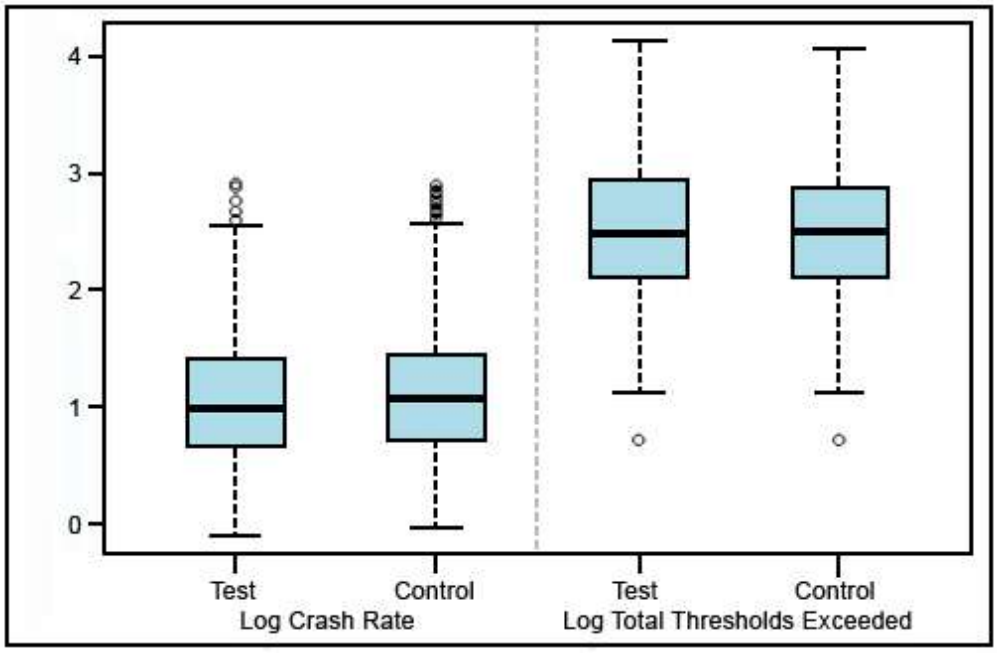

Figure 33. Warning Letter—Investigation—CSA

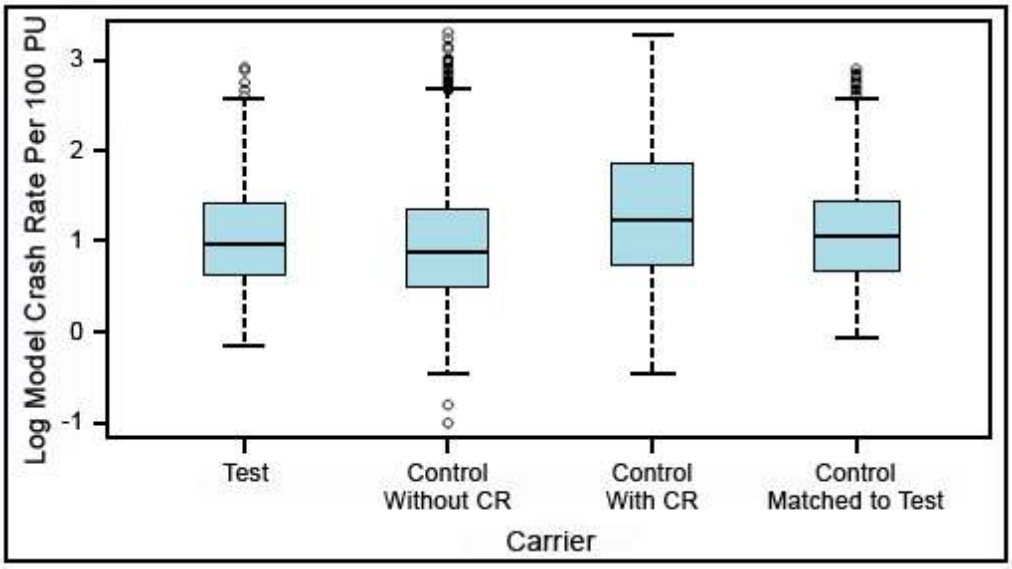

Figure 34. Warning Letter—Investigation—CSA

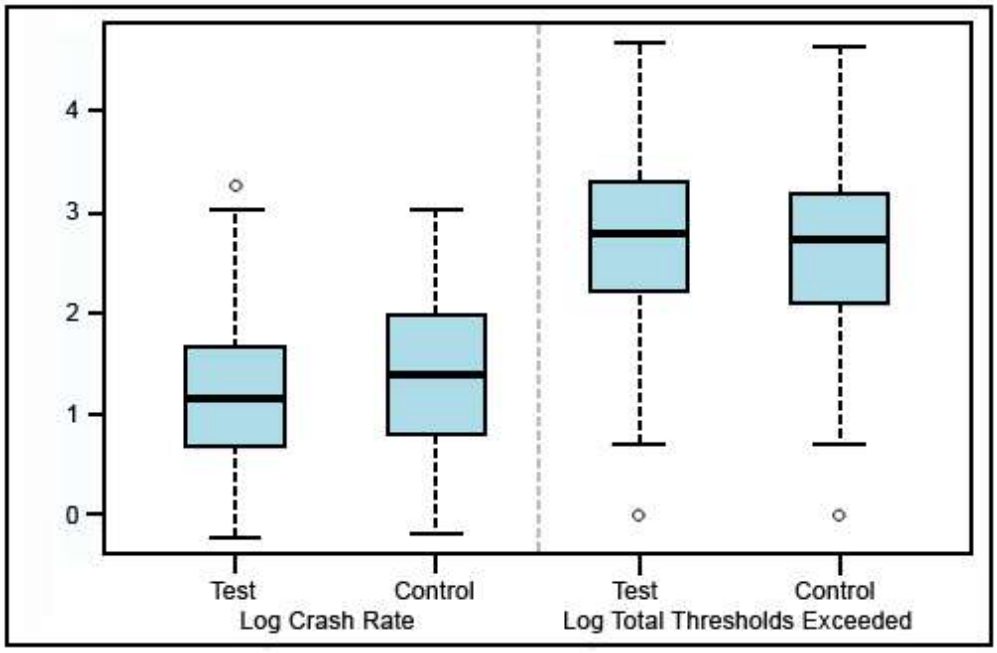

Figure 35. Interventions Ending in NOC

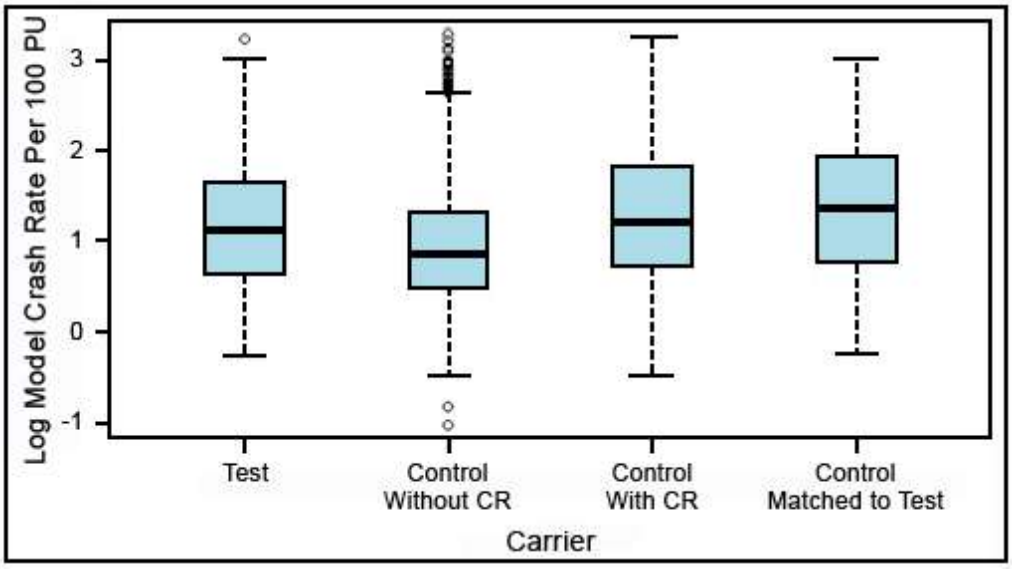

Figure 36. Interventions Ending in NOC

[This page intentionally left blank.]

APPENDIX D: SUPPORTING DATA FOR SAFESTAT A/B ANALYSIS (SECTION 9)

Table 104. Percent Exceeding At Least Two BASIC Thresholds Test

Month	At Least Two Thresholds Exceeded	Total Carriers	Percent
0	496	818	60.6%
1	462	781	59.2%
2	423	734	57.6%
3	391	702	55.7%
4	355	670	53.0%
5	334	641	52.1%
6	322	620	51.9%
7	297	593	50.1%
8	279	564	49.5%
9	239	543	44.0%
10	231	518	44.6%
11	194	483	40.2%
12	176	457	38.5%
13	163	416	39.2%
14	154	390	39.5%
15	135	357	37.8%
16	110	317	34.7%
17	94	289	32.5%
18	86	244	35.2%

Table 105. Percent Exceeding At Least Two BASIC Thresholds Test Control With CR

Month	At Least Two Thresholds Exceeded	Total Carriers	Percent
0	341	606	56.3%
1	331	590	56.1%
2	326	567	57.5%
3	303	548	55.3%
4	284	524	54.2%
5	261	499	52.3%
6	247	478	51.7%
7	230	462	49.8%
8	212	436	48.6%
9	199	416	47.8%
10	183	390	46.9%
11	160	356	44.9%
12	151	337	44.8%
13	136	310	43.9%
14	118	279	42.3%
15	97	240	40.4%
16	86	204	42.2%
17	71	179	39.7%
18	59	157	37.6%

Table 106. Percent Exceeding Unsafe Driving or Crash Indicator Thresholds Test

Month	Exceeded Unsafe Driving or Crash Indicator	Total Carriers	Percent
0	270	818	33.0%
1	255	781	32.7%
2	236	734	32.2%
3	218	702	31.1%
4	192	670	28.7%
5	180	641	28.1%
6	159	620	25.6%
7	150	593	25.3%
8	133	564	23.6%
9	120	543	22.1%
10	114	518	22.0%
11	96	483	19.9%
12	85	457	18.6%
13	80	416	19.2%
14	78	390	20.0%
15	69	357	19.3%
16	52	317	16.4%
17	52	289	18.0%
18	48	244	19.7%

Table 107. Percent Exceeding Unsafe Driving or Crash Indicator Thresholds Test Control With CR

Month	Exceeded Unsafe Driving or Crash Indicator	Total Carriers	Percent
0	225	606	37.1%
1	220	590	37.3%
2	215	567	37.9%
3	203	548	37.0%
4	182	524	34.7%
5	171	499	34.3%
6	156	478	32.6%
7	149	462	32.3%
8	134	436	30.7%
9	126	416	30.3%
10	110	390	28.2%
11	92	356	25.8%
12	79	337	23.4%
13	73	310	23.5%
14	62	279	22.2%
15	51	240	21.3%
16	45	204	22.1%
17	39	179	21.8%
18	37	157	23.6%

Table 108. Average Number of BASIC Thresholds Exceeded Test

Month	Total Carriers	Average Number BASICs Exceeded
0	818	1.81
1	781	1.78
2	734	1.76
3	702	1.71
4	670	1.63
5	641	1.59
6	620	1.55
7	593	1.52
8	564	1.45
9	543	1.36
10	518	1.37
11	483	1.28
12	457	1.23
13	416	1.20
14	390	1.22
15	357	1.18
16	317	1.12
17	289	1.05
18	244	1.10

Table 109. Average Number of BASIC Thresholds Exceeded Test Control With CR

Month	Total Carriers	Average Number BASICs Exceeded
0	606	1.80
1	590	1.77
2	567	1.78
3	548	1.73
4	524	1.73
5	499	1.69
6	478	1.66
7	462	1.61
8	436	1.56
9	416	1.52
10	390	1.50
11	356	1.50
12	337	1.44
13	310	1.43
14	279	1.43
15	240	1.38
16	204	1.47
17	179	1.44
18	157	1.36

Table 110. Average Log Crash Rate Per 100 Power Units Predicted by Model Test

Month	Total Carriers	Average Log Crash Rate Per 100 PU
0	817	1.59
1	772	1.60
2	714	1.60
3	668	1.60
4	628	1.59
5	593	1.59
6	567	1.58
7	537	1.58
8	500	1.57
9	477	1.54
10	449	1.57
11	415	1.53
12	392	1.50
13	352	1.52
14	327	1.50
15	300	1.48
16	262	1.43
17	234	1.41
18	197	1.45

Table 111. Average Log Crash Rate Per 100 Power Units Predicted by Model Test Control With CR

Month	Total Carriers	Average Log Crash Rate Per 100 PU
0	606	1.63
1	587	1.63
2	560	1.65
3	535	1.66
4	508	1.66
5	482	1.65
6	458	1.65
7	441	1.64
8	412	1.64
9	390	1.61
10	364	1.60
11	331	1.62
12	313	1.59
13	288	1.59
14	263	1.59
15	225	1.59
16	194	1.56
17	167	1.56
18	144	1.54

APPENDIX E: AN EMPIRICAL BAYES MODEL FOR PREDICTION OF CRASH RATES

The empirical Bayes model, including the data used based on a sampling procedure, was described at a high level in subsection 8.7. The model is now described in detail for the interested reader. The model requires specification of a likelihood for the data and a prior distribution for the parameters. The likelihood follows a Poisson distribution and the Poisson mean parameter follows a Gamma distribution. The Poisson–Gamma mixture results in a Negative Binomial likelihood for the crash counts that is used to calculate maximum likelihood estimates for parameters in a regression model. The following notation is used.

y – total number of crashes for a carrier over the 18-month period.
t – number of power units for a carrier.
λ – the Poisson rate parameter.
μ – the Gamma rate parameter.
β – Negative Binomial regression parameters.
x – predictor variables in regression function.
α – the Negative Binomial Scale Parameter.

Figure 37. Notations

Likelihood: For each carrier *i*, crashes are assumed to follow a Poisson distribution with rate lambda.

$$p(y_i \mid t_i, \lambda_i) = \frac{e^{-t_i\lambda_i}(t_i\lambda_i)^{y_i}}{y_i!}$$

Figure 38. Poisson Distribution

Prior: The rate parameter follows a Gamma distribution with mean mu and scale parameter alpha.

$$p(\lambda_i \mid \alpha, \mu_i) = \frac{1}{(\alpha\mu_i)^{1/\alpha}\,\Gamma(1/\alpha)}\ \lambda_i^{(1/\alpha)-1}\, e^{-\lambda_i/(\alpha\mu_i)}$$

Figure 39. Gamma Distribution

Negative Binomial: Combining likelihood with prior gives the Negative Binomial for *y*. The parameter mu is attached to a log-linear regression function.

$$p(y_i \mid \alpha, \mu_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(1/\alpha)\, y_i!}\left(\frac{1}{1 + 1/(\alpha t_i \mu_i)}\right)^{y_i}\left(\frac{1/(\alpha t_i \mu_i)}{1 + 1/(\alpha t_i \mu_i)}\right)^{1/\alpha}$$

$$\log \mu_i = \beta_0 + \beta_1 x_{i1} + \cdots\cdots + \beta_p x_{ip}$$

Figure 40. Negative Binomial

The Negative Binomial likelihood is maximized with respect to beta and alpha to give maximum likelihood estimates.

Empirical Bayes estimator: a weighted average of the observed rate *y/t* and the Negative Binomial mean mu.

$$E[\lambda_i \mid \alpha, \mu_i, y_i] = \frac{y_i + 1/\alpha}{t_i + 1/(\alpha\mu_i)} = \left(\frac{t_i\,\alpha\mu_i}{1 + t_i\,\alpha\mu_i}\right)\frac{y_i}{t_i} + \left(\frac{1}{1 + t_i\,\alpha\mu_i}\right)\mu_i$$

Figure 41. Empirical Bayes Estimator

The estimates of alpha and mu are used in the equation above to give the empirical Bayes estimator.

APPENDIX F: SURVEY FOR FIELD PERSONNEL

Thank you for participating in the CSA 2010 Operational Model Test Survey. Being actively involved in the daily operations of this test program, your responses are extremely important and will provide valuable feedback about the feasibility, practicality, and effectiveness of the CSA 2010 program. The goal is to find out which intervention processes worked well for FMCSA and which did not. When replying to the questions, please be as specific and concise as possible.

1. **We would like you to evaluate CSA2010's monthly intervention selection and carrier assignment process, and let us know what appears to be working well, what isn't working, and what needs to be changed. Please address the following items:**

   A. **CSA2010's prioritization of the carriers selected for intervention.**

      [use this space for your answer]

   B. **CSA2010's recommended type of intervention for the motor carrier.**

      [use this space for your answer]

   C. **Have there been times when you have had to override the recommended prioritization for the carrier? If yes, please discuss below (if not already discussed in your response to 1A or 1B, above). How often does this occur?**

      [use this space for your answer]

   D. **Have there been times when you have had to override the recommended type of intervention? If yes, please discuss below (if not already discussed in your response to 1A or 1B, above). How often does this occur?**

2. **Compared to SafeStat, how would you rate CSA2010's monthly intervention selection and carrier assignment process in terms of identifying carriers that are unsafe?**

[use this space for your answer]

3. **Considering the document request procedures, how often is the initial request for documents successful? Partially successful? Not successful at all? Can the task of document acquisition be improved?**

[use this space for your answer]

4. **What are the major problems, if any, in processing the documentation received from the carriers? How can these be addressed?**

[use this space for your answer]

5. **Consider the task of identifying motor carrier "Process Breakdowns" during the investigation. Please evaluate how well this works. What are the primary lessons learned in effectively and accurately identifying Process Breakdowns?**

[use this space for your answer]

6. **Is the procedure for determining the appropriate post-investigation intervention effective? Can it be improved?**

[use this space for your answer]

7. **Based on your experience with the Op Model Test, do you think that NOVs issued as a result of a CSA2010 investigation are an effective intervention tool? Please discuss.**

[use this space for your answer]

8. **Based on your experience with the Op Model Test, do you think that CSPs issued as a result of a CSA2010 investigation are an effective intervention tool? Please discuss.**

[use this space for your answer]

REFERENCES

1. Motor Carrier Safety Status Measurement System Methodology, Version 8.6, January 2004, John A. Volpe National Transportation Systems Center, Cambridge, MA. http://ai.fmcsa.dot.gov/CarrierResearchResults/PDFs/SafeStat_method.pdf.
2. Green, P.E., Matteson, A. 2007. Evaluation of 2006 Georgia Crash Data Reported to the MCMIS Crash File, UMTRI-2007-48.
3. Blower, D., Matteson, A. 2006. Evaluation of Missouri 2005 Crash Data Reported to the MCMIS Crash File. UMTRI-2006-32.
4. Green, P.E., Blower, D. 2005. Evaluation of New Jersey Crash Data Reported to the MCMIS Crash File. UMTRI-2005-13.
5. Green, P.E., Blower, D. 2008. Updated Ratio of Crash Severities Reportable to the MCMIS Crash File. UMTRI-2008-52.
6. Safety Measurement System (SMS) Methodology, Version 1.2, April 2009. John A. Volpe National Transportation Systems Center, Cambridge, MA.
7. Zaloshnja, E., Miller, T. 2006. Unit Costs of Medium and Heavy Truck Crashes, December 1006. Pacific Institute for Research and Evaluation, Calverton, MD.
8. Blower, D., Matteson, A. 2003. Evaluation of the Motor Carrier Management Information System Crash File, Phase One. UMTRI-2003-6.
9. Blower, D., Matteson, A. 2003. Patterns of MCMIS Crash File Underreporting in Ohio. UMTRI-2003-27.
10. Blower, D., Matteson, A. 2010. Evaluation of 2008 Kansas Crash Data Reported to the MCMIS Crash File. UMTRI-2010-5.
11. Green, P.E., Matteson, A. 2010. Evaluation of 2008 Florida Crash Data Reported to the MCMIS Crash File. UMTRI-2010-26.

[This page intentionally left blank.]