United States Government Accountability Office

GAO

Report to Congressional Requesters

June 2007

MOTOR CARRIER SAFETY

A Statistical Approach Will Better Identify Commercial Carriers That Pose High Crash Risks Than Does the Current Federal Approach

GAO

Accountability * Integrity * Reliability

GAO-07-585

GAO

Accountability·Integrity·Reliability

Highlights

Highlights of GAO-07-585, a report to congressional requesters

Why GAO Did This Study

The Federal Motor Carrier Safety Administration (FMCSA) has the primary federal responsibility for reducing crashes involving large trucks and buses that operate in interstate commerce. FMCSA decides which motor carriers to review for compliance with its safety regulations primarily by using an automated, data-driven analysis model called SafeStat. SafeStat uses data on crashes and other data to assign carriers priorities for compliance reviews.

GAO assessed (1) the extent to which changes to the SafeStat model could improve its ability to identify carriers that pose high crash risks and (2) how the quality of the data used affects SafeStat's performance. To carry out its work, GAO analyzed how SafeStat identified high-risk carriers in 2004 and compared these results with crash data through 2005.

What GAO Recommends

GAO is recommending that FMCSA use a negative binomial regression model to identify carriers that pose high crash risks.

In commenting on a draft of this report, the Department of Transportation agreed that the use of a negative binomial regression model looked promising for selecting carriers for compliance reviews, but expressed some reservation about the greater sensitivity of this approach to problems with reported crash data.

www.gao.gov/cgi-bin/getrpt?GAO-07-585.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Sidney H. Schwartz at (202) 512-7387 or schwartzsh@gao.gov, or Susan A. Fleming at (202) 512-2834 or flemings@gao.gov.

June 2007

MOTOR CARRIER SAFETY

A Statistical Approach Will Better Identify Commercial Carriers That Pose High Crash Risks Than Does the Current Federal Approach

What GAO Found

While SafeStat does a better job of identifying motor carriers that pose high crash risks than does a random selection, regression models GAO applied do an even better job. SafeStat works about twice as well as (about 83 percent better than) selecting carriers randomly. SafeStat is built on a number of expert judgments rather than using statistical approaches, such as a regression model. For example, its designers decided to weight more recent motor carrier crashes twice as much as less recent ones on the premise that more recent crashes were stronger indicators of future crashes. GAO estimates that if FMCSA used a negative binomial regression model, FMCSA could increase its ability to identify high-risk carriers by about 9 percent over SafeStat. Carriers identified by the negative binomial regression model as posing a high crash risk experienced 9,500 more crashes than those identified by the SafeStat model over an 18 month follow-up period. The primary use of SafeStat is to identify and prioritize carriers for FMCSA and state compliance reviews. FMCSA measures the ability of SafeStat to perform this role by comparing the crash rate of carriers identified as posing a high crash risk with the crash rate of other carriers. Using a negative binomial regression model would further FMCSA's mission of reducing crashes through the more effective targeting of compliance reviews to the set of carriers that pose the greatest crash risk.

Late-reported, incomplete, and inaccurate data reported to FMCSA by states have been a long-standing problem. However, GAO found that late reported data had a small effect on SafeStat's ability to identify carriers that pose high crash risks in 2004. If states had reported all crash data within 90 days after occurrence, as required by FMCSA, a net increase of 299 carriers (or 6 percent) would have been identified as posing high crash risks of the 4,989 that FMCSA identified. Reporting timeliness has improved, from 32 percent of crashes reported on time in fiscal year 2000, to 89 percent in fiscal year 2006. Regarding completeness, GAO found that data for about 21 percent of the crashes (about 39,000 of 184,000) exhibited problems that hampered linking crashes to motor carriers. Having complete information on crashes is important because SafeStat treats crashes as the most important factor for assessing motor carrier crash risk, and crash information is also the crucial factor in the statistical approaches that we employed. Regarding accuracy, a series of studies by the University of Michigan Transportation Research Institute covering 14 states found incorrect reporting of crash data is widespread. GAO was not able to quantify the effect of the incomplete or inaccurate data on SafeStat's ability to identify carriers that pose high crash risks because it would have required gathering crash records at the state level—an effort that was impractical for GAO. FMCSA has acted to improve crash data quality by completing a comprehensive plan for data quality improvement, implementing an approach to correct inaccurate data, and providing grants to states for improving data quality, among other things.

Contents

Figures

Abbreviations

FMCSA	Federal Motor Carrier Safety Administration
MCMIS	Motor Carrier Management Information System
SafeStat	Motor Carrier Safety Status Measurement System

United States Government Accountability Office
Washington, DC 20548

June 11, 2007

The Honorable James L. Oberstar
Chairman
The Honorable John L. Mica
Ranking Republican Member
Committee on Transportation and Infrastructure
House of Representatives

The Honorable Peter A. DeFazio
Chairman
The Honorable John J. Duncan
Ranking Republican Member
Subcommittee on Highways and Transit
Committee on Transportation and Infrastructure
House of Representatives

The Honorable Thomas E. Petri
House of Representatives

The Federal Motor Carrier Safety Administration (FMCSA) within the U.S. Department of Transportation has the primary federal responsibility for reducing crashes, deaths, and injuries involving large trucks and buses operating in interstate commerce. While it carries out a number of activities toward this end, an important tool at its disposal is the compliance review—a detailed inspection of a motor carrier's operations at its place of business. FMCSA decides which carriers to inspect primarily by using an automated, data-driven analysis system called the Motor Carrier Safety Status Measurement System (SafeStat). SafeStat uses data on crashes, vehicle and driver violations, and other information to develop numerical scores for carriers, and then SafeStat assigns each carrier a priority to receive a compliance review.

Following an incident in which a bus company, with many driver violations and a low priority for compliance review from the SafeStat model, suffered a fire on one of its buses that resulted in 23 deaths, you were interested in whether SafeStat could better identify commercial motor carriers at risk for crashes. To address your interest, we assessed (1) the extent to which changes to the SafeStat model could improve its ability to identify these carriers and (2) how the quality of the data used affects SafeStat's performance. These two topics are the main focus of this report. We also examined the findings of other studies on how SafeStat's ability to identify carriers at risk for crashes can be improved. (See app. I.)

To determine whether statistical approaches could be used to improve FMCSA's ability to identify carriers that pose high crash risks, we tested a number of regression models and compared their performance with SafeStat's results from June 2004. We chose 2004 because it allowed us to examine actual crash data for the 18-month period following June 2004 to determine the degree to which SafeStat successfully identified carriers that proved to be of high risk for crashes. It also allowed us to include crashes that occurred within the 18 months after June 2004 but had not yet been reported to FMCSA by December 2005. Using regression models, we compared the predictive performance of these statistical approaches to SafeStat's performance to determine which method best identified carriers that pose high crash risks. We also calculated crash rates from a series of random samples of all carriers to determine if the SafeStat model did a better job than random selection in identifying motor carriers that pose high crash risks. To assess whether changes could be made to the SafeStat model to improve its identification of carriers that pose high crash risks, we tested changes to selected portions of the SafeStat model and investigated the effect of changing decision rules used to construct the four safety evaluation areas.[1]

To assess the extent to which data quality affects SafeStat's ability to identify carriers that pose high crash risks, we carried out a series of analyses and surveyed the literature to identify findings from other studies. To address timeliness, we measured the number of days it took states to report crashes. We also added late-reported crashes to FMCSA's June 2004 data and recalculated SafeStat scores to determine the effect of late-reported crashes on carriers' rankings. For completeness, we attempted to match crash records in FMCSA's Motor Carrier Management Information System (MCMIS) crash master file to motor carriers listed in the MCMIS census file and reviewed studies on state reporting. To address accuracy, we reviewed a report that tested the accuracy of electronic data on a sample of paper records and studies that identified the impact of incorrectly reported crashes in individual states on MCMIS data quality. While there are known problems with the quality of the crash data

[1]There are four safety evaluation areas—accident, driver, vehicle, and safety management. They are used by the SafeStat model to assess a carrier's safety. See the background section for a description of these four areas. SafeStat is built on a number of expert judgments rather than using statistical approaches, such as a regression model.

reported to FMCSA for use in SafeStat, we determined that the data were of sufficient quality for our use, which was to compare the ability of regression models to identify carriers that pose high crash risks to the current approach, which is largely derived through professional judgment. We conducted our work in accordance with generally accepted government auditing standards from May 2006 through May 2007. Appendix II provides further information on our scope and methodology.

Shortly, we expect to issue a related report that examines how FMCSA identifies and takes action against carriers that are egregious safety violators. In addition, that report examines how thoroughly and consistently FMCSA conducts compliance reviews.

Results in Brief

While SafeStat does a better job of identifying motor carriers that pose high crash risks than does a random selection, regression models we applied do an even better job. SafeStat works about twice as well as (about 83 percent better than) selecting carriers randomly and, therefore, has value for improving safety. SafeStat is built on a number of expert judgments. For example, SafeStat's designers used their judgment and experience to weight more recent crashes involving a motor carrier twice as much as less recent crashes on the premise that more recent crashes were stronger indicators that a carrier may have crashes in the future. Using similar reasoning, fatal crashes were weighted more heavily than less serious crashes. We found that if a negative binomial regression model was used instead, FMCSA could increase its ability to identify carriers that pose high crash risks by about 9 percent over SafeStat.[2] Moreover, according to our analysis, this 9 percent improvement would enable FMCSA to identify carriers with twice as many crashes in the following 18 months as those carriers identified under its current approach.[3] Carriers identified by the negative binomial regression model as posing a high crash risk experienced 9,500 more crashes than those identified by the SafeStat model over an 18 month follow-up period. The primary use of SafeStat is to identify and prioritize carriers for FMCSA and state safety compliance reviews. FMCSA measures the ability of SafeStat to perform this role by comparing the crash rate of carriers identified as posing a high

[2]Negative binomial regression is often used to model count data (e.g., crashes). The results from this regression model can be interpreted as the estimated mean number of crashes per carrier.

[3]The 9 percent improvement is in crash rate per 1,000 vehicles over an 18-month period.

crash risk with the crash rate of other carriers. In our view, using a negative binomial regression model would further FMCSA's mission of reducing crashes through the more effective targeting of safety improvement and enforcement programs to the set of carriers that pose the greatest crash risk. Applying a regression model would be easy to adapt to the existing SafeStat model and, in our opinion, would be beneficial even if FMCSA makes major revisions to its compliance and enforcement approach in the coming years under its Comprehensive Safety Analysis 2010 initiative.[4]

Crash data reported by the states from December 2001 through June 2004 have problems in terms of timeliness, accuracy, and completeness that potentially hinder FMCSA's ability to identify high risk carriers. Regarding timeliness, we found that including late-reported data had a small impact on SafeStat—including late-reported data added a net of 299 (or 6 percent) more carriers to the original 4,989 carriers that the SafeStat model ranked as highest risk in June 2004.[5] The timeliness of crash reporting has shown steady and marked improvement: the percentage of crashes reported by states within 90 days of occurrence jumped from 32 percent in fiscal year 2000 to 89 percent in fiscal year 2006. Regarding completeness, data for about 21 percent of the crashes (about 39,000 of 184,000) exhibited problems that hampered linking crashes to motor carriers. Thirteen percent of the crashes (about 24,000) involving interstate carriers reported to FMCSA from December 2001 through June 2004 are missing the unique identifier that FMCSA assigns to each carrier when the agency authorizes the carrier to engage in interstate commerce. Crashes without a unique identifier to link to a company are excluded from use in SafeStat. An additional 8 percent of the crashes (about 15,000) that were reported had an identification number that could not be matched to a motor carrier in the FMCSA database that contains census information on motor carriers. Linking crashes to carriers is important because the current SafeStat

[4]The goal of this initiative is to develop an optimal operational model that will allow FMCSA to focus its resources on improving the safety performance of high-risk operators.

[5]We applied the SafeStat model to retrospective data. Because of changes to the MCMIS crash file over the past 2 years, our number does not correspond exactly to the number of carriers identified by FMCSA as high risk on June 25, 2004. Had all crash data been reported within 90 days of when the crashes occurred, 182 of the carriers identified by SafeStat as highest risk would have been excluded (because other carriers had higher crash risks), and 481 carriers that were not originally designated as posing high crash risks would have scored high enough to be considered high risk, resulting in a net addition of 299 carriers.

model treats crashes as the most important factor in assessing motor carrier crash risk. Crash information is also the crucial factor in the regression models that we employed. Regarding accuracy, a series of University of Michigan Transportation Research Institute's reports on crash reporting shows that, among the 14 states studied, incorrect reporting of crash data is widespread. For example, in recent reports, the researchers found that, in 2005, Ohio incorrectly reported 1,094 (22 percent) of the 5,037 cases it reported, and Louisiana incorrectly reported 137 (5 percent) of the 2,699 cases it reported. In Ohio, most of the 1,094 crashes did not qualify because they did not meet the crash severity threshold.[6] We were not able to quantify the actual effect of the incomplete or inaccurate data on SafeStat's ability to identify carriers that pose high crash risks, because it would have required us to gather crash records at the state level—an effort that was impractical. FMCSA has acted to improve the quality of SafeStat's data by completing a comprehensive plan for data quality improvement, implementing an approach to correct inaccurate data, and providing grants to states for improving data quality, among other things. We could not quantify the effects of FMCSA's efforts to improve the completeness or accuracy of the data for the same reason as mentioned above.

This report contains a recommendation to the Secretary of Transportation aimed at applying a negative binomial regression model to the four SafeStat safety evaluation areas that would result in better identification of commercial motor carriers that pose high crash risks. Because FMCSA has initiated efforts to improve the quality of SafeStat's data, we are not making a recommendation in this area.

In commenting on a draft of this report, the department agreed that it would be reasonable to consider the use of the negative binomial regression model in order to better target compliance reviews to carriers posing high crash risks, but expressed some concerns about placing more emphasis on crash information and less on other factors, such as driver, vehicle, or safety management issues. In addition, FMCSA noted that, while it has devoted considerable efforts to improving the quality of crash

[6]A reportable crash is one that meets both a vehicle and a crash severity threshold. Generally, for a crash to be reported, it must involve a truck with a gross vehicle weight rating of over 10,000 pounds; a bus with seating for at least nine people, including the driver; or a vehicle displaying a hazardous materials placard. Reportable accidents involve a fatality, an injury requiring transport to a medical facility for immediate medical attention, or towing required because the vehicle sustained disabling damage.

data submitted by states, the negative binomial regression model is more sensitive than SafeStat to problems with the crash data.

Background

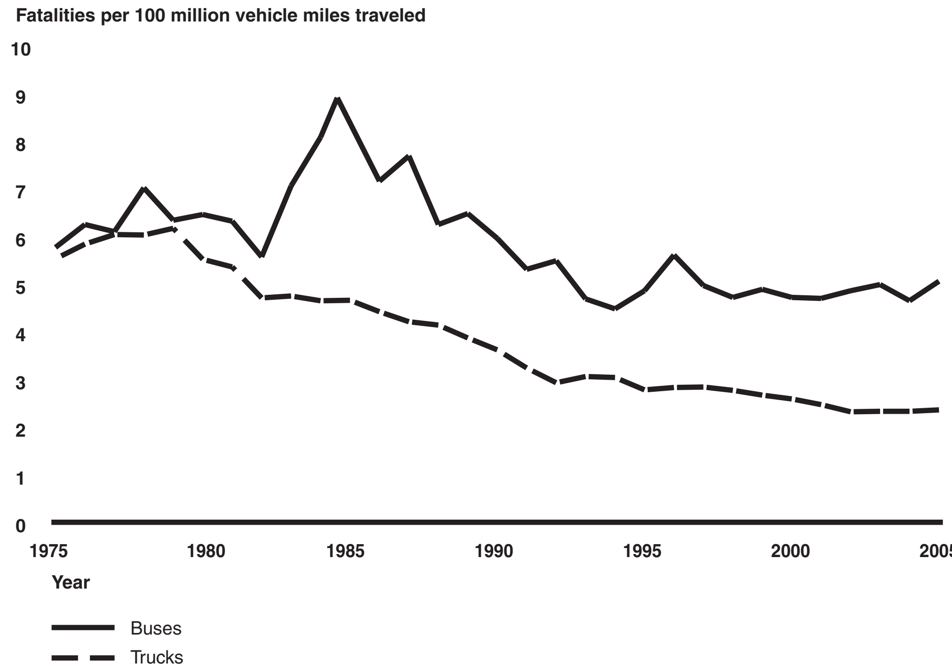
The interstate commercial motor carrier industry, primarily the trucking industry, is an important part of the nation's economy. Trucks transport over 11 billion tons of goods annually, or about 60 percent of the total domestic tonnage shipped.[7] Buses also play an important role, transporting an estimated 631 million passengers annually. There are approximately 711,000 commercial motor carriers registered in MCMIS,[8] about 9 million trucks and buses, and more than 10 million drivers. Most motor carriers are small; about 51 percent operate one vehicle, and another 31 percent operate two to four vehicles. Carrier operations vary widely in size, however, and some of the largest motor carriers operate upwards of 50,000 vehicles. Carriers continually enter and exit the industry. Since 1998, the industry has increased in size by an average of about 29,000 interstate carriers per year.

In the United States, commercial motor carriers account for less than 5 percent of all highway crashes, but these crashes result in about 13 percent of all highway deaths, or about 5,500 of the approximately 43,000 highway fatalities that occur nationwide annually. In addition, about 106,000 of the approximately 2.7 million highway injuries per year involve motor carriers. The fatality rate for trucks has generally been decreasing over the past 30 years, but this decrease has leveled off, and the rate has been fairly stable since the mid-1990s. The fatality rate for buses has improved slightly from 1975 to 2005 but has more annual variability than the fatality rate for trucks due to a much smaller total vehicle miles traveled. (See fig. 1.)

[7]This figure is for 2002, the most recent date for which data is available.

[8]This includes an unidentified number of carriers that are registered but are no longer in business.

Figure 1: Commercial Motor Vehicle Fatality Rate, 1975 to 2005



Source: GAO analysis of Department of Transportation data.

Congress created FMCSA through the Motor Carrier Safety Improvement Act of 1999 to reduce crashes, injuries, and fatalities involving commercial motor vehicles. To accomplish this mission, FMCSA carries out a number of enforcement, education, and outreach activities. FMCSA uses enforcement as its primary approach for reducing the number of crashes, fatalities, and injuries involving trucks and buses. Some of FMCSA's enforcement programs include compliance reviews, which are on-site reviews of carriers' records and operations to determine compliance with regulations; safety audits of new interstate carriers; and roadside inspections of drivers and vehicles.

FMCSA's education and outreach programs are intended to promote motor carrier safety and consumer awareness. One of the programs is the New Entrant program, which is designed to inform newly registered motor carriers about motor carrier safety standards and regulations to help them comply with FMCSA's requirements. Other programs are designed to identify unregistered carriers and get them to register, promote increased safety belt use among commercial drivers, and inform organizations and individuals that hire buses how to make safe choices. FMCSA plans to make major revisions to its compliance and enforcement approach under an initiative called Comprehensive Safety Analysis 2010.

Compliance reviews are an important enforcement tool because they allow FMCSA to take an in-depth look at carriers that have been identified as posing high crash risks because of high crash rates or poor safety performance records. Motor carriers may be identified as high risk from SafeStat or through calls to FMCSA's complaint hotline. Carriers are given a satisfactory, conditional, or unsatisfactory safety rating. A conditional rating means the carrier is allowed to continue operating, but FMCSA may schedule a follow-up compliance review to ensure that problems noted in the first compliance review are addressed. An unsatisfactory rating must be addressed or the carrier is placed out of service, meaning it is no longer allowed to do business, and the carrier may face legal enforcement actions undertaken by FMCSA. Compliance reviews can take several days to complete, depending on the size of the carrier, and may result in enforcement actions being taken against a carrier.

FMCSA uses both its own inspectors and state inspectors to carry out its enforcement activities. In total, about 750 staff are available to perform compliance reviews, and more than 10,000 staff do vehicle and driver inspections at weigh stations and other points. Together, FMCSA and its state partners perform about 16,000 compliance reviews a year, which cover about 2 percent of the nation's 711,000 carriers.[9]

Because the number of inspectors is small compared with the size of the motor carrier industry, FMCSA prioritizes carriers for compliance reviews. To do so, it uses SafeStat to identify carriers that pose high crash risks. SafeStat is a model that uses information gathered from crashes, roadside inspections, traffic violations, compliance reviews, and enforcement cases to determine a motor carrier's safety performance relative to that of other motor carriers that have similar exposure in these areas. A carrier's score is calculated on the basis of its performance in four safety evaluation areas:

- *Accident safety evaluation area:* The accident safety evaluation area reflects a carrier's crash history relative to other motor carriers' histories. The safety evaluation area is based on state-reported crash data, vehicle

[9]FMCSA completed 15,626 compliance reviews in 2006. The number of companies reviewed was less because some carriers received more than 1 compliance review.

data from MCMIS, and data on reportable crashes and annual vehicle miles traveled from the most recent compliance review. A carrier must have two or more reportable crashes within the last 30 months to have the potential to receive a deficient value and thus be made a priority for a compliance review.

- *Driver safety evaluation area:* The driver safety evaluation area reflects a carrier's driver-related safety performance and compliance relative to other motor carriers. The driver safety evaluation area is based on violations cited in roadside inspections that have been performed within the last 30 months and compliance reviews that have occurred within the last 18 months, together with the number of drivers listed in MCMIS. A carrier must have three or more driver inspections, three or more moving violations, or at least one acute or critical violation of driver regulations[10] from a compliance review to have the potential to receive a deficient value and thus be made a priority for a compliance review.

- *Vehicle safety evaluation area:* The vehicle safety evaluation area reflects a carrier's vehicle-related safety performance and compliance relative to other motor carriers. The vehicle safety evaluation area is based on violations identified during vehicle roadside inspections that have occurred within the last 30 months or vehicle-related acute and critical violations of regulations discovered during compliance reviews that have occurred within the last 18 months. A carrier must have either three or more vehicle inspections or at least one acute or critical violation of vehicle regulations from a compliance review to have the potential to receive a deficient value and thus be made a priority for a compliance review.

- *Safety management safety evaluation area:* The safety management safety evaluation area reflects a carrier's safety management relative to other motor carriers. It is based on the results of violations cited in closed enforcement cases in the past 6 years or violations of regulations related to hazardous materials and safety management discovered during a compliance review performed within the last 18 months. A carrier must have had at least one enforcement case initiated and closed or at least two

[10]Acute violations are violations so severe that FMCSA requires immediate corrective actions by a motor carrier regardless of the carrier's overall safety status. An example of an acute violation is a carrier's failing to implement an alcohol or drug testing program for drivers. Critical violations are serious, but less severe than acute violations, and most often point to gaps in carriers' management or operational controls. For example, a carrier may not maintain records of driver medical certificates.

enforcement cases closed within the past 6 years, or at least one acute, critical, or severe violation of hazardous material or safety management regulations[11] identified during a compliance review within the last 18 months to have the potential to receive a deficient value and thus be made a priority for a compliance review.

A motor carrier's score is based on its relative ranking, indicated as a value, in each of the four safety evaluation areas. For example, if a carrier receives a value of 75 in the accident safety evaluation area, then 75 percent of all carriers with sufficient data for evaluation performed better in that safety evaluation area, while 25 percent performed worse. The calculation used to determine a motor carrier's SafeStat score is as follows:

SafeStat Score = (2.0x accident value) + (1.5x driver value)

+ vehicle value + safety management value

As shown in the formula, the accident and driver safety evaluation areas have 2.0 and 1.5 times the weight, respectively, of the vehicle and safety management safety evaluation areas. Safety evaluation area values less than 75 are ignored in the formula used to determine the SafeStat score. For example, a carrier with values of 74 for all four safety evaluation areas has a SafeStat score of 0. FMCSA assigned more weight to these safety evaluation areas because, according to FMCSA, crashes and driver violations correlate relatively better with future crash risk. In addition, more weight is assigned to fatal crashes and to crashes that occurred within the last 18 months. In consultation with state transportation officials, insurance industry representatives, safety advocates, and the motor carrier industry, FMCSA used its expert judgment and professional knowledge to assign these weights, rather than determining them through a statistical approach, such as regression modeling.

FMCSA assigns carriers categories ranging from A to H according to their performance in each of the safety evaluation areas. A carrier is considered

[11]Severe violations are violations of hazardous materials regulations. Level I violations require immediate corrective actions. An example of a level I violation is offering or accepting a hazardous material for transportation in an unauthorized vehicle. Level II violations indicate a breakdown in the management or operational controls of the facility. An example of a level II violation is failing to train hazardous materials employees as required.

to be deficient in a safety evaluation area if it receives a value of 75 or higher in that particular safety evaluation area. Although a carrier may receive a value in any of the four safety evaluation areas, the carrier receives a SafeStat score only if it is deficient in one or more safety evaluation areas. Carriers that are deficient in two or more safety evaluation areas and have a SafeStat score of 225 or more are considered to pose high crash risks and are placed in category A or B. (See table 1.) Carriers that are deficient in two safety evaluation areas but have a SafeStat score of less than 225 are placed in category C and receive a medium priority for compliance reviews. Carriers that are deficient in only one of the safety evaluation areas are placed in category D, E, F, or G. Carriers that are not deficient in any of the safety evaluation areas do not receive a SafeStat score and are placed in category H.

Table 1: SafeStat Categories and Their Priority for Compliance Reviews

Category	Condition	Priority for compliance review
A	Deficient in all four safety evaluation areas *or* Deficient in three safety evaluation areas that result in a weighted SafeStat score of 350 or more	High
B	Deficient in three safety evaluation areas that result in a weighted SafeStat score of less than 350 *or* Deficient in two safety evaluation areas that result in a weighted SafeStat score of 225 or more	High
C	Deficient in two safety evaluation areas that result in a weighted SafeStat score of less than 225	Medium
D	Deficient in the accident safety evaluation area (accident safety evaluation area value between 75-100)	Low
E	Deficient in the driver safety evaluation area (driver safety evaluation area value between 75-100)	Low
F	Deficient in the vehicle safety evaluation area (vehicle safety evaluation area value between 75-100)	Low
G	Deficient in the safety management safety evaluation area (safety management safety evaluation area value between 75-100)	Low
H	Not deficient in any of the safety evaluation areas (value below 75 in each of the safety evaluation areas)	Low

Source: GAO summary of FMCSA data.

Of the 622,000 motor carriers listed in MCMIS as having one or more vehicles in June 2004, about 140,000, or 23 percent, received a SafeStat category A through H. There are several reasons why a small proportion of carriers receive a score. First, approximately 305,900, or about 42 percent, of the carriers have crash, vehicle inspection, driver inspection, or enforcement data of any kind. SafeStat relies on these data to calculate a motor carrier's score, so carriers without such data are not rated by SafeStat. It is likely that some of the carriers listed in MCMIS are no longer in business, but it is also possible that these carriers had no crashes, inspections, or compliance reviews in the 30-month period prior to June 2004. Second, a carrier must meet the minimum requirements to be assigned a value in a given safety evaluation area.[12] If, for example, a carrier had only one reportable crash within the last 30 months, then the carrier would not be assigned an accident safety evaluation area value. Of the 305,900 carriers that have any safety data in SafeStat, 140,000 met the SafeStat minimum requirements in one or more safety evaluation areas. Of these 140,000 carriers, 45,000 were rated in categories A through G. The other carriers were placed in category H because they were not considered deficient, meaning they did not receive a value of 75 or more in any of the safety evaluation areas.

The design of SafeStat and its data sufficiency requirements increase the likelihood that larger motor carriers will be deficient in one of the safety evaluation areas, in other words, rated in categories A through G, than are small carriers. About 51 percent of all carriers listed in MCMIS operate one vehicle, and about 3 percent of them received a SafeStat rating in categories A through G. (See table 2.) In contrast, fewer than 1 percent of the carriers listed in MCMIS have more than 100 vehicles, and nearly 25 percent of them received a SafeStat rating in categories A through G.

[12]Minimum requirements in this context mean that the carrier has enough safety data to receive a rating. Usually, the safety data are associated with adverse safety events. However it is possible for a carrier to have enough roadside inspections, even if none of the inspections resulted in violations, to qualify for a driver and vehicle safety evaluation area score.

Table 2: Size Distribution of Carriers Receiving a SafeStat Rating of A through G

Carrier size (number of vehicles)	Number of carriers (percentage[a])	Number of carriers within size category receiving A through G SafeStat rating (percentage of carriers in size category)
1	317,037 (51%)	8,697 (3%)
>1 to 4	191,739 (31%)	14,430 (8%)
>4 to 10	66,422 (11%)	10,595 (16%)
>10 to 25	28,780 (5%)	6,504 (23%)
>25 to 100	14,148 (2%)	3,550 (25%)
>100	3,903 (1%)	909 (23%)

Source: GAO analysis of FMCSA data.

Note: The table only includes those carriers listed as having one or more vehicles.

[a]Percentages do not equal 100 because of rounding.

A Statistical Approach Would Better Identify Carriers That Pose High Crash Risks Than Does FMCSA's Current Approach

We found that FMCSA could improve SafeStat's ability to identify carriers that pose high crash risks if it applied a statistical approach, called a negative binomial regression model, to the four SafeStat safety evaluation areas instead of its current approach. Through this change, FMCSA could more efficiently target compliance reviews to the set of carriers that pose the greatest crash risk. Applying a negative binomial regression model would improve the identification of high risk carriers over SafeStat's performance by about 9 percent,[13] compared with the current approach, which incorporates safety data weighted in accordance with the professional judgment and experience of SafeStat's designers. Moreover, according to our analysis, this 9 percent improvement would enable FMCSA to identify carriers with almost twice as many crashes in the following 18 months as those carriers identified under its current approach. Targeting these high-risk carriers would result in FMCSA giving compliance reviews to carriers that experienced both a higher crash rate and, in conjunction with the higher crash rate, 9,500 more crashes over an 18-month period than those identified by the SafeStat model. Applying a negative binomial regression model approach to the SafeStat safety evaluation areas would be easy to implement and, in our opinion, would be consistent with other FMCSA uses for SafeStat beyond identifying

[13]The 9 percent improvement is in the crash rate per 1,000 vehicles over an 18-month period.

carriers that pose high risks for crashes. In addition, adopting a negative binomial regression model approach would be beneficial even if FMCSA makes major revisions to its compliance and enforcement approach in the coming years under its Comprehensive Safety Analysis 2010 initiative. Overall, other changes to the SafeStat model that we explored, such as modifying decision rules used in the construction of the safety evaluation areas, did not improve the model's overall performance.

Regression Models Identify Carriers That Pose High Crash Risks Better Than Expert Judgment

Although SafeStat is nearly twice as effective as (83 percent better than) random selection in identifying carriers that pose high crash risks[14] and, therefore, has value for improving safety, we found that FMCSA could improve SafeStat's ability to identify such carriers by about 9 percent if it applied a negative binomial regression model approach to its analysis of motor carrier safety data. The use of a regression model does not entail assigning the letter categories currently assigned by the SafeStat model. Rather, the model predicts carriers' crash risks, sorts the carriers according to their risk level, and assigns a high priority for a compliance review to the highest risk carriers. The improvement in identification of high-risk carriers, which we observed with the negative binomial regression model, is consistent with results obtained in an earlier analysis of MCMIS data performed by a team of researchers at Oak Ridge National Laboratory.[15]

To compare the effectiveness of regression models and SafeStat in identifying carriers that pose high crash risks, we applied several regression models to the four safety evaluation areas (accident, driver, vehicle, and safety management) used by the SafeStat model. We recalculated SafeStat's June 2004 accident safety evaluation area values because the data FMCSA provided on the number of crashes for each

[14]Applying the SafeStat model to June 2004 data identifies 4,989 carriers as high risk (categories A or B). Using 10,000 randomly selected samples of 4,989 carriers and considering the crashes that these carriers had between June 2004 and December 2005, we found that the crash rate per 1,000 vehicles in the ensuing 18 months was 83 percent higher among the carriers identified by the SafeStat model than among the randomly selected carriers.

[15]Ken Campbell, Rich Schmoyer, and Ho-Ling Hwang, *Review of the Motor Carrier Safety Status Measurement System (SAFESTAT)*, Oak Ridge National Laboratory, Final Report, October 2004. See appendix I for a more detailed discussion of the findings from this report.

carrier differed in 2006 from the data used in the model in 2004.[16] Using our accident safety evaluation area value and the original driver, vehicle, and safety management safety evaluation area values from June 2004, we selected the 4,989 carriers that our regression models identified as the highest crash risks,[17] calculated the crash rate per 1,000 vehicles for these carriers over the next 18 months, and compared this rate with the crash rate per 1,000 vehicles for the 4,989 carriers identified by the SafeStat model as posing high crash risks (categories A and B).

All of the regression models that we estimated were at least as effective as SafeStat in identifying motor carriers that posed high crash risks. (See app. III for these results.) Of these, the negative binomial regression approach gave the best results and proved 9 percent more effective than SafeStat, as measured by future crashes per 1,000 vehicles. The set of carriers in SafeStat categories A and B had a crash rate of 102 per 1,000 vehicles for the 18 months after June 2004 while the set of high-risk carriers identified by the negative binomial regression model had 111 crashes per 1,000 vehicles. Even though this 9 percent improvement rate seems modest, it translates into nearly twice as many "future crashes" identified. Specifically, the negative binomial regression model identified carriers that had nearly twice as many crashes (from July 2004 to December 2005) as the carriers identified by SafeStat—19,580 crashes compared with 10,076.[18]

SafeStat (categories A and B) and our negative binomial regression model identified many of the same carriers—1,924 of the 4,989 (39 percent)—as posing high crash risks. However, our model also identified a number of high-risk carriers that SafeStat did not identify, and vice versa. For example, our model identified 2,244 carriers as posing high crash risks, while SafeStat placed these carriers in category D (the accident area), assigning them a lower priority for compliance reviews. One reason for

[16]This occurs because data were added, deleted, or modified as more information became known over time. See appendix III for a more detailed discussion.

[17]The threshold could be increased or decreased to align with the resources that FMCSA and its state partners have available to perform compliance reviews. As discussed earlier, FMCSA and its state partners select carriers for these reviews because they pose high crash risks and for other reasons.

[18]The carriers identified as high risk by SafeStat had a total of 98,619 vehicles while those identified by the negative binomial regression model had 175,820 vehicles. The identification of larger sized companies on average by the negative binomial regression model is how a 9 percent increase in the crash rate translated into 9,500 additional crashes.

this difference is the decision rules that SafeStat employs. Under SafeStat, carriers must perform worse than 75 percent of all carriers to be considered deficient in any safety evaluation area. The regression approach identifies the carriers with the highest crash risks regardless of how they compare with their peers in individual areas. For example, we identified as posing high crash risks 482 carriers that SafeStat did not consider at all for compliance reviews because the carriers had not performed worse than 75 percent of their peers in any of the four safety evaluation areas.

FMCSA Can Apply a Regression Model Approach in the Short Term, Even Though It Is Planning to Overhaul SafeStat

In the short term, FMCSA could easily implement a regression model approach for SafeStat.[19] All the information required as input for the negative binomial regression model is already entered into SafeStat. In addition, a standard statistical package can be used to apply the negative binomial approach to the four SafeStat safety evaluation areas. Like SafeStat, the negative binomial regression model would be run every month to produce a list of motor carriers that pose high crash risks, and these carriers would then be assigned priorities for a compliance review. As with SafeStat, the results of the negative binomial model would change slightly each month with the addition of new safety data to MCMIS.

In discussing the concept of adopting a negative binomial regression model approach with FMCSA officials, they were interested in understanding how the use of the negative binomial regression model results could be used to identify and improve the safety of those carriers that pose the greatest crash risks (much as the SafeStat categories of A and B do now) and how it could employ the proposed approach for current uses beyond identifying carriers that pose high crash risks. These uses include providing an understandable public display to shippers, insurers, and others who are interested in the safety of carriers; selecting carriers for roadside inspections; and trying to gain carriers' compliance with driver and vehicle safety rules, when these carriers may not have crashes, consistent with agency efforts.

- *Identifying and improving the safety of carriers that pose high crash risks.* The negative binomial regression model approach would produce a rank order listing of carriers by crash risk and by the predicted number of

[19]FMCSA can use the current safety evaluation area values in SafeStat and the number of state-reported crashes for each carrier in the 30 preceding months in the negative binomial regression model.

crashes. For compliance reviews, FMCSA could choose those carriers with the greatest number of predicted crashes. FMCSA would choose the number of carriers to review based on the resources available to it, much as it currently does.

Regarding improving the safety of carriers that pose high crash risks, FMCSA currently enrolls carriers that receive a SafeStat category of A, B, or C in the Motor Carrier Safety Improvement Program. This program aims to improve the safety of high-risk carriers through (1) a repetitive cycle of identification, data gathering, and assessment and (2) progressively harsher treatments applied to carriers that do not improve their safety. The use of a negative binomial regression model would not affect the structure or workings of this program, other than to better identify carriers that pose high crash risks. As discussed above, FMCSA would use the regression model's results to identify the highest risk carriers and then intervene using its existing approaches (such as issuing warning letters, conducting follow-up compliance reviews, or levying civil penalties) as treatment.

- *Providing an understandable display to the public.* FMCSA could choose to provide a rank order listing of carriers together with the associated number of predicted crashes or it could look for natural breaks in the predicted number of crashes and associate a category—such as "category A" to these carriers.

- *Selecting carriers for roadside inspections.* Safety rankings from the SafeStat model are also used in FMCSA's Inspection Selection System to prioritize carriers for roadside driver and vehicle inspections. The negative binomial regression model optimizes the identification of carriers by crash risk using safety evaluation area information. The negative binomial regression model approach that we describe in this report retains SafeStat's basic design with four safety management areas (driver, vehicle, accident, and safety management). Therefore, FMCSA could use the negative binomial regression model results to identify carriers that pose a high crash risk, the results from the driver and vehicle safety evaluation areas, or both, to target carriers or vehicles for roadside driver and vehicle inspections.

- *Furthering agency efforts to gain compliance with driver and vehicle safety rules for carriers that do not experience crashes (or a sufficient number of crashes to pose a high risk for crashes).* FMCSA was interested in understanding how, if at all, the negative binomial regression model approach would affect gaining compliance against carriers that may routinely violate safety rules (such as drivers' hours of service requirements), but where these violations do not lead to crashes. As discussed above, the negative binomial regression model approach retains SafeStat's four safety evaluation areas. Where it differs, is that it assigns different weights to those areas based on a statistical procedure, rather than having the weights assigned by expert judgment. As a result, FMCSA would still be able to identify carriers with many driver, vehicle, and safety management violations.

Other opportunities also exist for FMCSA to improve the ability of regression models to identify carriers that pose high crash risks. In 2005, a FMCSA compliance review work group reported a positive correlation between driver hours of service violations and crash rates.[20] Because FMCSA can link violations of specific regulatory provisions, including those limiting driver hours of service, to the crash experience of the carriers involved, it has the opportunity to improve the violation severity weighting used in constructing the driver and vehicle safety evaluation areas. FMCSA has detailed violation data from roadside inspections and can statistically analyze these data to find other strong relationships with carriers' crash risks. Changes made to the safety evaluation area methodology to strengthen the association with crash risk will improve the ability of the negative binomial regression model to identify carriers that pose high crash risks.

FMCSA has expressed doubts in the past when analysts have proposed switching to a regression model approach. For example, Oak Ridge National Laboratory advocated using a regression model approach in place of SafeStat in 2004, but FMCSA was reluctant to move away from its expert judgment model because it believed that the regression model approach would place undue weight on the accident safety evaluation area in determining priorities for compliance reviews,[21] thereby diminishing the incentive for motor carriers to comply with the many safety regulations that feed into the driver, vehicle, and safety management safety evaluation

[20]Federal Motor Carrier Safety Administration Compliance Review Workgroup, *Phase II Final Report: Proposed Operational Model for FMCSA Compliance and Safety Programs Report*, February 2005.

[21]Oak Ridge National Laboratory statistically measured the weights for the safety evaluation areas and estimated the accident safety evaluation area should have a weight of 57 in the SafeStat model formula. This compares with the present weight of 2 that SafeStat gives the accident safety evaluation area. Ken Campbell, Rich Schmoyer, and Ho-Ling Hwang, *Review of the Motor Carrier Safety Status Measurement System (SAFESTAT)*, Oak Ridge National Laboratory, Final Report, October 2004.

areas. In FMCSA's view, carriers would be less likely to comply with these regulations because violations in the driver, vehicle, and safety management areas would be less likely to lead to compliance reviews under a regression model approach that placed a heavy emphasis on crashes. Our view is that adopting a negative binomial regression model approach would better identify carriers that pose high crash risks and would thus further FMCSA's primary mission of ensuring safe operating practices among commercial interstate motor carriers.

Over the longer term, FMCSA is considering a complete overhaul of its safety fitness determinations with its Comprehensive Safety Analysis 2010 initiative. This planned comprehensive review and analysis of the agency's compliance and enforcement programs may result in a new operational model for identifying drivers and carriers that pose safety problems and for intervening to address those problems.[22] FMCSA expects to deploy the results of this initiative in 2010. In our opinion, given the relative ease of adopting the regression modeling approach discussed in this report,[23] and the immediate benefits that can be achieved, there is no reason to wait for FMCSA to complete its initiative, even if the initiative results in major revisions to the SafeStat model.

Modifications of SafeStat Did Not Improve Crash Identification

Besides investigating whether the use of regression models could improve SafeStat's ability to identify carriers that pose high crash risks, we explored whether the existing model could be improved by changing several of its decision rules. Overall, these changes did not enhance the model's ability to identify carriers that pose high crash risks. As long as FMCSA continues to estimate the safety evaluation area values with its present methodology, the rules we investigated help make the identification of high-risk motor carriers more efficient for both SafeStat and the negative binomial regression model.

Because the SafeStat model is composed of many components, we selected three decision rules for analysis. We chose these three rules because they are important pillars of the SafeStat model's methodology for constructing the safety evaluation areas and because we could complete

[22]We expect to issue a report shortly that provides additional discussion of FMCSA's initiative to identify and take action against carriers that are egregious safety violators.

[23]Revisions to SafeStat are exempt from notice and comment under the Administrative Procedure Act if they relate to FMCSA's internal practices and procedures.

our analysis of them during the time we had to perform our work. A fuller exploration of areas with high potential to improve the identification of carriers that pose high crash risks would be a long-term effort, and FMCSA plans to address this work as part of the Comprehensive Safety Analysis 2010 initiative.

- *Removing comparison groups.* As part of its methodology for calculating the accident, driver, and vehicle safety evaluation area values, SafeStat divides carriers into comparison groups. For example, in the driver safety evaluation area, SafeStat groups carriers by the number of moving violations they have, placing them in one of four groups (3 to 9, 10 to 28, 29 to 94, and 95 or more).[24] SafeStat uses the comparison groups to control for the size of the carrier. We removed all the comparison groups in each of the three safety evaluation areas, recalculated their values, and compared the number of crashes in which the carriers were involved and their crash rates, for each of the SafeStat categories A through H, with the SafeStat results in which comparison groups were retained.

- *Removing minimum event requirements.* SafeStat imposes minimum event requirements. For example, as noted, SafeStat does not consider a carrier's moving violations if, in the aggregate, its drivers had fewer than three moving violations over a 30-month period. FMCSA does not calculate a safety evaluation area value for carriers with fewer than three events in an attempt to control for carriers that have infrequent, rather than possibly systemic, safety problems.[25] We removed the requirement to have a minimum number of events (such as moving violations, crashes, and inspections), recalculated the three safety evaluation values, and compared the number of crashes in which the carriers were involved and their crash rates, for each of the SafeStat categories A through H, with the SafeStat results in which minimum event requirements were retained.

[24]SafeStat does not consider carriers with fewer than three moving violations.

[25]Carriers with one or zero state-reported crashes do not receive an accident safety evaluation area score unless the recordable accident indicator is available from a recent compliance review. Carriers with two or fewer driver inspections and two or fewer moving violations do not receive a driver safety evaluation area score unless the driver review indicator is available from a recent compliance review. Carriers with two or fewer vehicle inspections do not receive a vehicle safety evaluation area score unless the vehicle review indicator is available from a recent compliance review. In the data we reviewed, almost 2 percent of the carriers had undergone a compliance review within the 18 months prior to the SafeStat run on June 25, 2004.

- *Removing time and severity weights.* The SafeStat formula weights more recent events and more severe events more heavily than less recent or less severe events in the accident, driver, and vehicle safety evaluation areas. For example, the results of vehicle roadside inspections performed within the latest 6 months receive three times the weight of inspections performed 2 years ago. Similarly, crashes involving deaths or injuries receive twice as much weight as those that resulted in property damage only. We removed the time and severity weights for the three safety evaluation areas, recalculated these values, and compared the number of crashes in which the carriers were involved and their crash rates, for each of the SafeStat categories A through H, with the SafeStat results in which time and severity weights were retained.

- *Simultaneous changes to comparison group, event, and time severity requirements.* Finally, we simultaneously removed comparison groups, minimum event requirements, and time and severity weights and compared the number of crashes in which the carriers were involved and their crash rates, for each of the SafeStat categories A through H, with the SafeStat results in which comparison groups, minimum event requirements, and time and severity weights were retained.

The results of each of our individual analyses and of making all changes simultaneously produced one of two outcomes, neither of which was considered more desirable. Relaxing the minimum data requirements greatly increased the number of carriers identified as high risk without increasing the overall number of predicted crashes over the subsequent 18 months, thus reducing the effectiveness of the SafeStat model. Removing comparison groups and removing time and severity weights had the effect of reducing the future crashes per 1,000 vehicles among those carriers identified as high risk, also reducing the effectiveness of the SafeStat model. As a result, we are not reporting on these results in detail. Trying to modify the decision rules used in SafeStat did highlight the balance that FMCSA has to strike between maximizing the identification of companies with the largest number of crashes (usually larger carriers) and those carriers with the greatest safety risk (which can be of any size).

Despite Quality Problems, FMCSA's Crash Data Can Be Used to Compare Methods for Identifying Carriers That Pose High Crash Risks

The quality of crash data is a long-standing problem that potentially hindered FMCSA's ability to accurately identify carriers that pose high crash risks.[26] Despite the problems of late-reported crashes and incomplete and inaccurate data on crashes during the period we studied, we determined that the data were of sufficient quality for our use, which was to assess how the application of regression models might improve the ability to identify high-risk carriers over the current approach—not to determine absolute measures of crash risk. Our reasoning is based on the fact that we used the same data set to compare the results of the SafeStat model and the regression models. Limitations in the data would apply equally to both results. FMCSA has recently undertaken a number of efforts to improve crash data quality.

Late Reporting Had a Small Effect on SafeStat's Ability to Identify High-Risk Carriers

FMCSA's guidance provides that states report all crashes to MCMIS within 90 days of their occurrence. Late reporting can cause SafeStat to miss some of the carriers that should have received a SafeStat score. Alternatively, since SafeStat's scoring involves a relative ranking of carriers, a carrier may receive a SafeStat score and have to undergo a compliance review because crash data for a higher risk carrier were reported late and not included in the calculation.

Late reporting affected SafeStat's ability to identify all high-risk carriers to a small degree—about 6 percent—for the period that we studied. Late reporting of crashes by states affected the safety rankings of more than 600 carriers, both positively and negatively. When SafeStat analyzed the 2004 data, which did not include the late-reported crashes, it identified 4,989 motor carriers as highest risk, meaning they received a category A or B ranking. With the addition of late-reported crashes, 481 carriers moved into the highest risk category, and 182 carriers dropped out of the highest risk category, resulting in a net increase of 299 carriers (6 percent) in the highest risk category. After the late-reported crashes were added, 481 carriers that originally received a category C, D, E, F, or G SafeStat rating received an A or B rating. These carriers would not originally have been given a high priority for a compliance review because the SafeStat calculation did not take into account all of their crashes. On the other hand, a small number of carriers would have received a lower priority if

[26]For another assessment of data quality, see Office of Inspector General, *Improvements Needed in the Motor Carrier Safety Status Measurement System*, U.S. Department of Transportation, Report MH-2004-034, 2004.

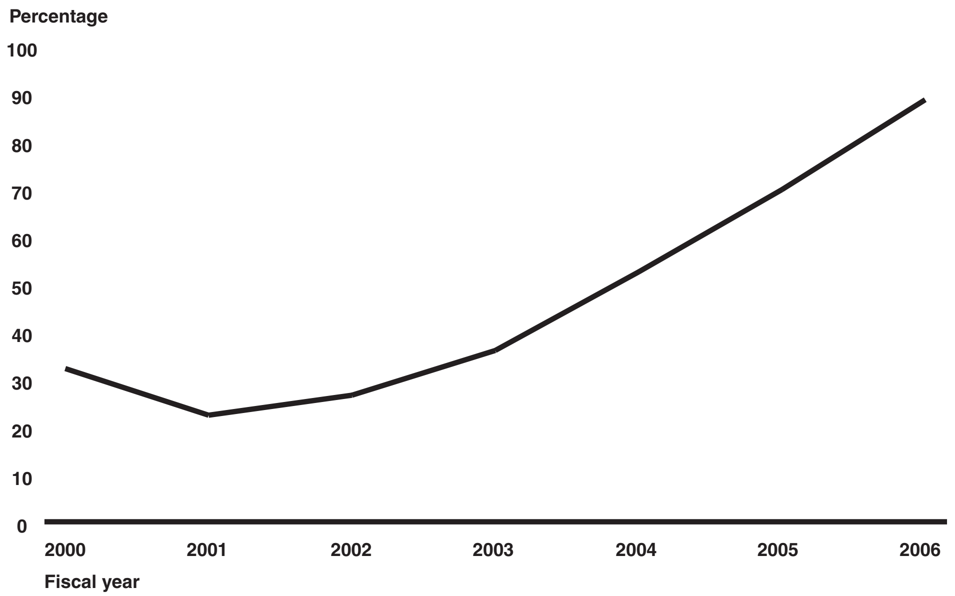
the late-reported crashes had been included in their score. Specifically, 182 carriers – or fewer than 4 percent of those ranked, fell from the A or B category into the C, D, E, F, or G category once the late-reported crashes were included.[27] These carriers would not have been considered high priority for a compliance review if all crashes had been reported on time. This does not have a big effect on the overall motor carrier population, however, as only 4 percent of carriers originally identified as highest risk were negatively affected by late reporting.

The timeliness of crash reporting has shown steady and marked improvement. The median number of days it took states to report crashes to MCMIS dropped from 225 days in calendar year 2001 to 57 days in 2005 (the latest data available at the time of our analysis).[28] In addition, the percentage of crashes reported by states within 90 days of occurrence has jumped from 32 percent in fiscal year 2000 to 89 percent in fiscal year 2006. (See fig. 2.)

[27]These 182 carriers were no longer in the worst 25 percent for the accident safety evaluation area after the addition of the late-reported crashes.

[28]Part of the improvement in timeliness of reporting for the most recent year is that an unknown number of crashes that occurred in 2005 had still not been reported as of June 2006, the date we obtained these data.

Figure 2: Percentage of Crashes Submitted to MCMIS within 90 Days of Occurrence



Source: GAO analysis of FMCSA data.

Incomplete Data from States Potentially Limit SafeStat's Identification of All Carriers That Pose High Crash Risks

FMCSA uses a motor carrier identification number, which is unique to each carrier, as the primary means of linking inspections, crashes, and compliance reviews to motor carriers. Approximately 184,000 (76 percent) of the 244,000 crashes reported to MCMIS between December 2001 and June 2004 involved interstate carriers. Of these 184,000 crashes, nearly 24,000 (13 percent) were missing this identification number. As a result, FMCSA could not match these crashes to motor carriers or use them in SafeStat. In addition, the carrier identification number could not be matched to a number listed in MCMIS for 15,000 (8 percent) other crashes that involved interstate carriers. Missing data or being unable to match data for nearly one quarter of the crashes during the period of our review potentially has a large impact on a motor carrier's SafeStat score because SafeStat treats crashes as the most important source of information for assessing motor carrier crash risk. Theoretically, information exists to match crash records to motor carriers by other means, but such matching would require too much manual work to be practicable.

We were not able to quantify the actual effect of either the missing data or the data that could not be matched for MCMIS overall. To do so would have required us to gather crash records at the state level—an effort that was impractical. For the same reason, we cannot quantify the effects of FMCSA's efforts to improve the completeness of the data (discussed later). However, a series of reports by the University of Michigan Transportation Research Institute sheds some light on the completeness of the data submitted to MCMIS by the states.[29] One of the goals of the research was to determine the states' crash reporting rates. Reporting rates varied greatly among the 14 states studied, ranging from 9 percent in New Mexico in 2003 to 87 percent in Nebraska in 2005. It is not possible to draw wide-scale conclusions about whether state reporting rates are improving over time because only two of the states—Missouri and Ohio—-were studied in multiple years. However, in these two states, the reporting rate did improve. Missouri experienced a large improvement in its reporting rate, with 61 percent of eligible crashes reported in 2001, and 83 percent reported in 2005. Ohio's improvement was more modest, increasing from 39 percent in 2000 to 43 percent in 2005.

The University of Michigan Transportation Research Institute's reports also identified a number of factors that may affect states' reporting rates. One of the main factors affecting reporting rates is the reporting officer's understanding of crash reporting requirements. The studies note that reporting rates are generally lower for less serious crashes and for crashes involving smaller vehicles, which may indicate that there is some confusion about which crashes are reportable. Some states, such as Missouri, aid the officer by explicitly listing reporting criteria on the police accident reporting form, while other states, such as Washington, leave it up to the officer to complete certain sections of the form if the crash is reportable, but the form includes no guidance on reportable crashes. Yet other states, such as North Carolina and Illinois, have taken this task out of officers' hands and include all reporting elements on the police accident reporting form. Reportable crashes are then selected centrally by the state, and the required data are transmitted to MCMIS.

[29]The University of Michigan Transportation Research Institute's reports on state crash reporting can be found at http://www.umtri.umich.edu. State reports issued by the University of Michigan Transportation Research Institute cover California, Florida, Illinois, Iowa, Louisiana, Maryland, Michigan, Missouri, Nebraska, New Jersey, New Mexico, North Carolina, Ohio, and Washington. We included all of these reports in our review.

Inaccurate Data Potentially Limit SafeStat's Ability to Identify Carriers That Pose High Crash Risks

Inaccurate data, such as reporting a nonqualifying crash to FMCSA, potentially has a large impact on a motor carrier's SafeStat score because SafeStat treats crashes as the most important source of information for assessing motor carrier crash risk. For the same reasons as discussed in the preceding section, we were neither able to quantify these effects nor determine how data accuracy has improved for MCMIS overall.

The University of Michigan Transportation Research Institute's reports on crash reporting show that, among the 14 states studied, incorrect reporting of crash data is widespread. In recent reports, the researchers found that, in 2005, Ohio incorrectly reported 1,094 (22 percent) of the 5,037 cases, and Louisiana incorrectly reported 137 (5 percent) of the 2,699 cases. In Ohio, most of the incorrectly reported crashes did not qualify because they did not meet the crash severity threshold. In contrast, most of the incorrectly reported crashes in Louisiana did not qualify because they did not involve vehicles eligible for reporting. Other states studied by the institute had similar problems with reporting crashes that did not meet the criteria for reporting to MCMIS. These additional crashes could cause some carriers to exceed the minimum number of crashes required to receive a SafeStat rating and result in SafeStat's mistakenly identifying carriers as posing high crash risks. Because each report focuses on reporting in one state in a particular year, it is not possible to identify the number of cases that have been incorrectly reported nationwide and, therefore, it is not possible to determine the impact of inaccurate reporting on SafeStat's calculations.

As noted in the University of Michigan Transportation Research Institute's reports, states may be unintentionally submitting incorrect data to MCMIS because of difficulties in determining whether a crash meets the reporting criteria. For example, in Missouri, pickups are systematically excluded from MCMIS crash reporting, which may cause the state to miss reportable crashes. However, some pickups may have vehicle weights above the reporting threshold, making crashes involving them eligible for reporting. There is no way for the state to determine which crashes involving pickups qualify for reporting without examining the characteristics of each vehicle. In this case, the number of omissions is likely to be relatively small, but this example demonstrates the difficulty states may face when identifying reportable crashes.

In addition, in some states, the information contained in the police accident report may not be sufficient for the state to determine if a crash meets the accident severity threshold. It is generally straightforward to determine whether a fatality occurred as a result of a crash, but it may be difficult to determine whether an injured person was transported for medical attention or a vehicle was towed because of disabling damage. In some states, such as Illinois and New Jersey, an officer can indicate on the form if a vehicle was towed by checking a box, but there is no way to identify whether the reason for towing was disabling damage. It is likely that such uncertainty results in overreporting because some vehicles may be towed for other reasons.

FMCSA Has Undertaken Efforts to Improve Crash Data Quality

FMCSA has taken steps to try and improve the quality of crash data reporting. As we noted in November 2005, FMCSA has undertaken two major efforts to help states improve the quality of crash data.[30] One program, the Safety Data Improvement Program, has provided funding to states to implement or expand activities designed to improve the completeness, timeliness, accuracy, and consistency of their data. FMCSA has also used a data quality rating system to rate and display ratings for states' crash and inspection data quality. Due to its public nature, this map serves as an incentive for states to make improvements in their data quality.

To further improve these programs, FMCSA has made additional grants available to states and implemented our recommendations to (1) establish specific guidelines for assessing states' requests for funding to support data improvement in order to better assess and prioritize the requests and (2) increase the usefulness of its state data quality map as a tool for monitoring and measuring commercial motor vehicle crash data by ensuring that the map adequately reflects the condition of the states' commercial motor vehicle crash data.

In February 2004, FMCSA implemented Data Q's, an online system that allows for challenging and correcting erroneous crash or inspection data. Users of this system include motor carriers, the general public, state officials, and FMCSA. In addition, in response to a recent recommendation by the Department of Transportation Inspector General, FMCSA is planning to conduct a number of evaluations of the effectiveness of a training course on crash data collection that it will be providing to states by September 2008.

[30]GAO, *Highway Safety: Further Opportunities Exist to Improve Data on Crashes Involving Commercial Motor Vehicles*, GAO-06-102 (Washington, D.C.: Nov. 18, 2005).

While the quality of crash reporting is sufficient for use in identifying motor carriers that pose high crash risks and has started to improve, commercial motor vehicle crash data continue to have some problems with timeliness, completeness, and accuracy. These problems have been well-documented in several studies, and FMCSA is taking steps to address the problems through studies of each state's crash reporting system and grants to states to fund improvements. As a result, we are not making any recommendations in this area.

Conclusion

Interstate commerce involving large trucks and buses has been growing substantially, and this growth is expected to continue. While the number of fatalities per million vehicle miles traveled has generally decreased over the last 30 years, the fatality rate has leveled off and remained fairly steady since the mid-1990s. FMCSA could more effectively address fatalities due to crashes involving a commercial motor vehicle if it better targeted compliance reviews to those carriers that pose the greatest crash risks. Using a negative binomial regression model would further FMCSA's mission of reducing crashes through the more effective targeting of compliance reviews to the set of carriers that pose the greatest crash risks. In light of possible changes to FMCSA's safety fitness determinations resulting from its Comprehensive Safety Analysis 2010 initiative, we are not suggesting that FMCSA undertake a complete and thorough investigation of SafeStat. Rather, we are advocating that FMCSA apply a statistical approach that employs the negative binomial regression model rather than relying on the current SafeStat formula that was determined through expert judgment. In our view, the substitution of a statistically based approach would likely yield a markedly better ability to identify carriers that pose high crash risks with relatively little time or effort on FMCSA's part.

Recommendation for Executive Action

We recommend that the Secretary of Transportation direct the Administrator of FMCSA to apply a negative binomial regression model, such as the one discussed in this report, to enhance the current SafeStat methodology.

Agency Comments and Our Evaluation

We provided a draft of this report to the Department of Transportation for its review and comment. In response, departmental officials, including FMCSA's Director of the Office of Enforcement and Compliance and Director of the Office of Research and Analysis, noted that our report provided useful insights and offered a potential avenue for further improving the effectiveness of FMCSA's efforts to reduce crashes involving motor carriers. The agency indicated that it is already working to improve upon SafeStat as part of its Comprehensive Safety Analysis 2010 initiative. FMCSA agreed that it would be useful for it to consider whether there are both short and longer term measures that would incorporate the type of analysis identified in our report, as an adjunct to the SafeStat model, in order to better target compliance reviews so as to make the best use of FMCSA's resources to reduce crashes.

The agency expressed some concerns with the negative binomial regression analysis, noting that its intent is to effectively target its compliance activities based on a broader range of factors than is considered in the negative binomial regression analysis approach described in our draft report, which increases reliance on past crashes as a predictor of future crashes while apparently de-emphasizing known driver, vehicle, or safety management compliance issues. FMCSA told us that it incorporates a broad range of information including driver behavior, vehicle condition, and safety management in an attempt to capture and enable the agency to act on accident precursors in order to reduce crashes.

FMCSA is correct in concluding that the use of the negative binomial regression approach could tilt enforcement heavily toward carriers that have experienced crashes and away from other aspects of its problem areas, such as violation of vehicle safety standards, that are intended to prevent crashes. That is because the present SafeStat model does not statistically assign weights to the accident, driver, vehicle, and safety management areas. In addition, the negative binomial regression approach fully considers information on the results of driver and vehicle inspection data and safety management data. We used the same data that FMCSA used, with some adjustments as new information became available. While we found that the driver, vehicle, and safety management evaluation area scores are correlated with the future crash risk of a carrier, the accident evaluation area correlates the most with future crash risk. We recognize that FMCSA selects carriers for compliance reviews for multiple reasons, such as to respond to complaints, and we would expect that it would retain this flexibility if it adopted the negative binomial regression approach.

FMCSA also indicated that greater reliance on crash data increases emphasis on the least reliable available data set, and one that is out of the organization's direct control—crash reporting. While our draft report found that crash reporting has improved, and that late reporting has not significantly impaired FMCSA's use of the SafeStat model, FMCSA noted that the reliance on previous crashes in the negative binomial regression analysis described in our draft report could result in greater sensitivity to the crash data quality issues.

As FMCSA noted in its comments, our results showed that the effect of late-reported data was minimal. Also, as mentioned in our draft report and in this final report, it was not practical to determine the effect, if any, on SafeStat rankings of correcting inaccurate data or adding incomplete data. Since June 2004, FMCSA has devoted considerable efforts to improving the quality of the crash data it receives from the states. States are now tracked quarterly for the completeness, timeliness, and accuracy of their crash reporting. As FMCSA continues its efforts to have states improve these data, any sensitivity of results to crash data quality issues for the negative binomial regression approach should diminish.

We are sending copies of this report to congressional committees and subcommittees with responsibility for surface transportation safety issues; the Secretary of Transportation; the Administrator, FMCSA; and the Director, Office of Management and Budget. We also will make copies available to others upon request. In addition, this report will be available at no charge on the GAO Web site at http://www.gao.gov.

If you have any questions about this report, please either contact Sidney H. Schwartz at (202) 512-7387 or Susan A. Fleming at (202) 512-2834. Alternatively, they may be reached at schwartzsh@gao.gov or flemings@gao.gov. Contact points for our Offices of Congressional

Relations and Public Affairs may be found on the last page of this report. Staff who made key contributions to this report are Carl Barden, Elizabeth Eisenstadt, Laurie Hamilton, Lisa Mirel, Stephanie Purcell, and James Ratzenberger.

Sidney H. Schwartz
Director
Applied Research and Methods

Susan A. Fleming
Director
Physical Infrastructure Issues

Appendix I: Results of Other Assessments of the SafeStat Model's Ability to Identify Motor Carriers That Pose High Crash Risks

Several studies by the Volpe National Transportation Systems Center (Volpe), the Department of Transportation's Office of Inspector General, the Oak Ridge National Laboratory (Oak Ridge), and others have assessed the predictive capability of the Motor Carrier Safety Status Measurement System (SafeStat) model and the data used by that model. In general, those studies that assessed the predictive power of SafeStat offered suggestions to increase that power, and those studies that assessed data quality found weaknesses in the data that the Federal Motor Carrier Safety Administration (FMCSA) relies upon.

Assessments of SafeStat's Predictive Capability

The studies we reviewed covered topics such as comparing SafeStat with random selection to determine which does a better job of selecting carriers that pose high crash risks, assessing whether statistical approaches could improve that selection, and analyzing whether carrier financial positions or driver convictions are associated with crash risk.

Predictive Capability of SafeStat Compared with Random Selection

In studies of the SafeStat model published in 2004 and 1998,[1] Volpe analyzed retrospective data to determine how many crashes the carriers in SafeStat categories A and B experienced over the following 18 months. The 2004 study used the carrier rankings generated by the SafeStat model on March 24, 2001. Volpe then compared the SafeStat carrier safety ratings with state-reported data on crashes that occurred between March 25, 2001, and September 24, 2002, to assess the model's performance. For each carrier, Volpe calculated a total number of crashes, weighted for time and severity, and then estimated a crash rate per 1,000 vehicles for comparing carriers in SafeStat categories A and B with the carriers in other SafeStat categories. The 1998 Volpe study used a similar methodology. Each study used a constrained subset of carriers rather than the full list contained in the Motor Carrier Management Information System (MCMIS).[2] Both studies found that the crash rate for the carriers in SafeStat categories A

[1]David Madsen and Donald Wright, Volpe National Transportation Systems Center, *An Effectiveness Analysis of SafeStat (Motor Carrier Safety Status Measurement System)*, Paper No. 990448, November 1998 and John A. Volpe National Transportation Systems Center, Motor Carrier Safety Assessment Division, *SafeStat Effectiveness Study Update*, March 2004.

[2]Volpe included only carriers with two or more crashes and/or three or more inspections during the preceding 30 months, and/or an enforcement action within the past 6 years, and/or a compliance review within the previous 18 months. This is consistent with the SafeStat minimum event requirements.

and B was substantially higher than the other carriers during the 18 months after the respective SafeStat run. On the basis of this finding, Volpe concluded that the SafeStat model worked.

In response to a recommendation by the Department of Transportation's Office of Inspector General,[3] FMCSA contracted with Oak Ridge to independently review the SafeStat model. Oak Ridge assessed the SafeStat model's performance and used the same data set (for March 24, 2001), provided by Volpe, that Volpe had used in its 2004 evaluation. Perhaps not surprisingly, Oak Ridge obtained a similar result for the weighted crash rate of carriers in SafeStat categories A and B over the 18-month follow-up period. As with the Volpe study, the Oak Ridge study was constrained because it was based on a limited data set rather than the entire MCMIS data set.

Application of Regression Models to Safety Data

While SafeStat does better than simple random selection in identifying carriers that pose high crash risks, other methods can also be used to achieve this outcome. Oak Ridge extended Volpe's analysis by applying regression models to identify carriers that pose high crash risks. Specifically, Oak Ridge applied a Poisson regression model and a negative binomial model using the safety evaluation area values as independent variables to a weighted count of crashes that occurred in the 30 months before March 24, 2001. (For more information on statistical analyses, see app. III.)

In addition, Oak Ridge applied the empirical Bayes method to the negative binomial regression model and assessed the variability of carrier crash counts by estimating confidence intervals. Oak Ridge found that the negative binomial model worked well in identifying carriers that pose high crash risks. However, the data set Oak Ridge used did not include any carriers with one reported crash in the 30 months before March 24, 2001. Because data included only carriers with zero or two or more reported crashes, the distribution of crashes was truncated.

Since the Oak Ridge regression model analysis did not cover carriers with safety evaluation area data and one reported crash, the findings from the study are limited in their generalizability. However, other analyses of crashes at intersections and on road segments have also found that the

[3]Office of Inspector General, *Improvements Needed*, 2004.

negative binomial regression model works well.[4] In addition, our analysis using a more recent and more comprehensive data set supports the finding that the negative binomial regression model performs better than the SafeStat model.

The studies carried out by other authors advocate the use of the empirical Bayes method in conjunction with a negative binomial regression model to estimate crash risk. Oak Ridge also applied this model to identify motor carriers that pose high crash risks. We applied this method to the 2004 SafeStat data and found that the empirical Bayes method best identified the carriers with the largest number of crashes in the 18 months after June 25, 2004. However, the crash rate per 1,000 vehicles was much lower than that for carriers in SafeStat categories A and B. We analyzed this result further and found that although the empirical Bayes method best identifies future crashes, it is not as effective as the SafeStat model or the negative binomial regression model in identifying carriers with the highest future crash rates. The carriers identified with the empirical Bayes method were invariably the largest carriers. This result is not especially useful from a regulatory perspective. Companies operating a large number of vehicles often have more crashes over a period of time than smaller companies. However, this does not mean that the larger company is necessarily violating more safety regulations or is less safe than the smaller company. For this reason, we do not advocate the use of the empirical Bayes method in conjunction with the negative binomial regression model as long as the method used to calculate the safety evaluation area values remains unchanged. If changes are made in how carriers are rated for safety, this method may in the future offer more promise than the negative binomial regression model alone.

Relationship of Carrier Financial Data and Safety Risk

Conducted on behalf of FMCSA, a study by Corsi, Barnard, and Gibney in 2002 examined how a carrier's financial performance data correlate with the carrier's score on a compliance review.[5] The authors selected those motor carriers from MCMIS in December 2000 that had complete data for

[4]Ezra Hauer, Douglas Harwood, and Michael Griffith, *The Empirical Bayes Method for Estimating Safety: A Tutorial.* Transportation Research Record 1784, National Academies Press, 2002, 126-131.

[5]Thomas Corsi, Richard Barnard, and James Gibney, *Motor Carrier Industry Profile: Linkages Between Financial and Safety Performance Among Carriers in Major Industry Segments*, Robert H. Smith School of Business at the University of Maryland, October 2002.

the accident, driver, vehicle, and safety management safety evaluation areas. Using these data, the authors then matched a total of 700 carriers to company financial statements in the annual report database of the American Trucking Associations.[6] The authors created a binary response variable for whether the carrier received a satisfactory or an unsatisfactory outcome on the compliance review. The authors then assessed how this result correlated with financial measures derived from the company financial statements. In general, the study found that indicators of poor financial condition correlated with an increased safety risk.

Two practical considerations limit the applicability of the findings from this study to SafeStat. First, the 700 carriers in the study sample are not necessarily representative of the motor carriers that FMCSA oversees. Only about 2 percent of the carriers evaluated by the SafeStat model in June 2004 had a value for the safety management safety evaluation area. Of these carriers, not all had complete data for the other three safety evaluation areas. Second, FMCSA does not receive annual financial statements from all motor carriers.[7] For these reasons, we did not consider using carrier financial data in our analysis of the SafeStat data.

Relationship of Commercial Driver License Convictions and Crash Risk

A series of studies by Lantz and others examined the effect of incorporating conviction data from the state-run commercial driver license data system into the calculation of a driver conviction measure.[8] The studies found that the driver conviction measure is weakly correlated with the crash per vehicle rate.[9] However, the studies did not incorporate the proposed driver conviction measure into one of the existing safety evaluation areas and use the updated measure to estimate new SafeStat

[6]The American Trucking Associations is a membership organization with a mission to serve and represent the interests of the trucking industry.

[7]The *Annual Report Form M* is required only for class 1 or class 2 carriers that have revenue exceeding $3 million for 3 consecutive years.

[8]Brenda Lantz and David Goettee, *An Analysis of Commercial Vehicle Driver Traffic Conviction Data to Identify Higher Safety Risk Motor Carriers*, Upper Great Plains Transportation Institute and FMCSA, 2004. Brenda Lantz, *Development and Implementation of a Driver Safety History Indicator into the Roadside Inspection Selection System*, FMCSA, April 2006.

[9]Correlation = 0.085. (See FMCSA, *Development and Implementation of a Driver Safety History Indicator into the Roadside Inspection Selection System*, April 2006, 14).

scores for carriers. While the use of commercial driver license conviction data may have potential for future incorporation into a model for identifying carriers that pose high crash risks, there is no assessment of its impact at this time.

Impact of Data Quality on SafeStat's Predictive Capability

The 2004 Office of Inspector General report, the 2004 Oak Ridge study, and reports by the University of Michigan Transportation Research Institute on state crash reporting all examined the impact of data quality on SafeStat's ability to identify carriers that pose high crash risks. These studies looked at issues such as late reporting and incomplete or inaccurate reporting of crash data and found weaknesses.

Late Reporting of Crash Data

To determine whether states promptly report SafeStat data, the Office of Inspector General conducted a two-stage statistical sample in which it selected 10 states for review and then selected crash and inspection reports from those states for examination. It sampled 392 crash records and 400 inspection records from July through December 2002. In 2 of the 10 states selected, Pennsylvania and New Mexico, no crash records were available for the sample period, so it selected samples from earlier periods. The Office of Inspector General also discussed reporting issues with state and FMCSA officials and obtained crash records from selected motor carriers. In addition, the Office of Inspector General used the coefficient of variation to analyze data consistency and trends in reporting timeliness across geographic regions.[10] Our review of the study indicates that it was based on sound audit methodology.

The study found that, as of November 2002, states submitted crash reports in fiscal year 2002 an average of 103 days after the crash occurred and that states varied widely in the timeliness of their crash data reporting. (FMCSA requires that states report crashes no more than 90 days after they occur.) In addition, the study found that 20 percent of the crashes that occurred in fiscal year 2002 were entered into MCMIS 6 months or more after the crash occurred. On the basis of this information, the Office of Inspector General concluded that the calculation of the accident safety

[10]The Office of Inspector General used MCMIS data to estimate a standard deviation for days to report a crash and then divided the standard deviation by the average number of days. This number was multiplied by 100 to derive the coefficient of variation. The obtained value of about 77 indicates substantial variability relative to the average number of days to report a crash.

evaluation area value was affected by the location of the carrier's operations but did not estimate the degree of this effect.

We also assessed the extent of late reporting. We measured how many days, on average, it took each state to report crashes to MCMIS in each calendar year and found that the amount of time taken to report crashes declined from 2000 to 2005. Our findings were similar in nature to the Office of Inspector General's findings. However, our results are broader because they are based on all crash data rather than a sample. In addition, since our work is more recent, it reflects more current conditions. We both came to the conclusion, although to varying degrees, that late reporting of crash data by states negatively affects SafeStat's identification of carriers that pose high crash risks.

Oak Ridge also examined the impact of late reporting. Using data provided by Volpe, Oak Ridge looked at the difference between the date a crash occurred and the date it was entered into MCMIS. The researchers found that after 497 days, 90 percent of the reported crashes were entered into MCMIS.

The Oak Ridge study also reran the SafeStat model for March 2001 with the addition of crash data from March 2003 to see how more complete data changed SafeStat scores. The study found that the addition of late-reported data increased the number of carriers in the high-risk group by 18 percent. This late reporting affected the rankings of 8 percent of all the carriers ranked by SafeStat in March 2001. Of these affected carriers, 3 percent moved to a lower SafeStat category, and 5 percent moved to a higher category. Including the late-reported crash data available in March 2003 for the period from September 1998 through March 2001 resulted in a 35 percent increase in the available crash data.

We performed the same analysis as the Oak Ridge study and obtained similar results. We used SafeStat data from June 2004, which include carrier safety data from December 2001 through June 2004. Using FMCSA's master crash file from June 2006, we found that, with the addition of late-reported crashes, 481 carriers moved into the highest risk category, and 182 carriers dropped out of the highest risk category resulting in a net increase of 299 carriers (6 percent) being added to the highest risk category.

The University of Michigan Transportation Research Institute issued a series of reports examining crash reporting rates in 14 states. These reports looked at late reporting as a potential source of low crash reporting rates but did not specifically examine the extent of late reporting or the impact of late reporting on SafeStat scores. The institute looked at reporting rates in each of the states by month to determine if reporting rates were lower in the latter part of the year because of late reporting. It found that reporting rates were lower in the latter part of the year in 6 of the 14 states studied. This issue was not a focus of our efforts, so we did not conduct a similar analysis.

Incomplete and Inaccurate Reporting of Crash Data

The Office of Inspector General's study found several instances of incomplete or inaccurate data on crashes and carriers. The study reviewed MCMIS reporting for all states and found that 6 of them did not report any crashes to FMCSA in the 6-month period from July through December 2002. In addition, the study found that MCMIS listed about 11 percent of carriers as having no vehicles, and 15 percent as having no drivers. Finally, from a sample of crash records, the study estimated that 13 percent of the crash reports and 7 percent of the inspection reports in MCMIS contained errors that would affect SafeStat results. In particular, the study concluded that the database identified the wrong motor carrier as having been involved in a crash or as having received a violation in 11 percent of the erroneous records.

The University of Michigan Transportation Research Institute also examined the accuracy of states' crash data reporting. To determine if crashes were reported accurately, the institute compared information contained in the individual states' police accident reporting files with crash data reported to MCMIS. Some states, such as Ohio, had enough information captured in the police accident file to determine if individual crashes were eligible for reporting, and, therefore, the institute was able to use these data in its analyses. In other states, not enough information was available to make a determination, and the institute had to project results on the basis of other states' experience. The institute also carried out a number of analyses, such as comparing reporting rates for different reporting jurisdictions, in an attempt to identify reporting trends in the individual states.

The institute identified several problems with the accuracy of states' crash reporting. All 14 states that it studied reported ineligible crashes to MCMIS. These crashes were ineligible because they either involved vehicles not eligible for reporting or they did not meet the crash severity threshold. In total, the 14 states reported nearly 5,800 ineligible crashes to MCMIS out of almost 68,000 crashes reported (9 percent). The states also failed to report a number of eligible crashes: the 14 states studied reported from 9 percent to 87 percent of eligible crashes.

Our review of the institute's methodology indicates that its findings are based on sound methodology and that its analyses were very thorough. However, its studies are limited to the 14 states studied and to the particular year studied. (Not all studies covered the same year.) These states' experience may or may not be representative of the experiences of the entire country, and there is no way to determine if the reporting for this year is representative of the state's reporting activities over a number of years or if the results were unique to that particular year. The exceptions to this are the studies for Missouri, which covered calendar years 2001 and 2005, and Ohio, which covered calendar years 2000 and 2005.

We did not attempt to assess the extent of inaccurate reporting in individual states, but we did find examples of inaccurate data reporting. To analyze the completeness of reporting, we attempted to match all crash records in the MCMIS master crash file for crashes occurring between December 26, 2001, and June 25, 2004, to the list of motor carriers in the MCMIS census file. We found that Department of Transportation numbers were missing for 30 percent of the crashes that were reported, and the number did not match a Department of Transportation number listed in MCMIS for 8 percent of reported crashes. We also compared the number of crashes in MCMIS with the number in the General Estimates System produced by the National Highway Traffic Safety Administration and found evidence of underreporting of crashes to MCMIS.[11]

[11]The General Estimates System collects all types of information from all types of crashes. It is based on a nationally representative probability sample from the estimated 6.4 million police-reported crashes that occur annually. While the crash eligibility definitions are not strictly comparable, the number of crashes reported to MCMIS is below the lower bound for the 95 percent confidence interval around the estimated total number of crashes for large trucks in 2004.

Appendix II: Scope and Methodology

To determine whether statistical approaches could be used to improve FMCSA's ability to identify carriers that pose high crash risks, we tested a variety of regression models and compared their results with results from the existing SafeStat model. The models we tested, using MCMIS data used by SafeStat in June 2004 to identify carriers that pose high crash risks, include the Poisson, negative binomial, zero-inflated negative binomial, zero-inflated Poisson, and empirical Bayes. We chose these regression models because crash totals for a company represent count outcomes, and these statistical models are appropriate for use with count data. In addition, we explored logistic regression to assess the odds of having a crash. Based on the results of the statistical models, we ranked the predicted means (or predicted probabilities in the logistic regression) to see which carriers would be at risk during the 18-month period after June 2004. We selected June 2004 because this date enabled us to examine MCMIS data on actual crashes that occurred in the 18-month period from July 2004 through December 2005.[1] We used these data to determine the degree to which SafeStat identified carriers that proved to pose high crash risks. We then compared the predictive performance of the regression models with the performance of SafeStat to determine which method best identified carriers that pose high crash risks. Using a series of simple random samples,[2] we also calculated the crash rates of all carriers listed in the main SafeStat summary results table in MCMIS for comparison with the crash rates of carriers identified by SafeStat as high risk. We did this analysis to determine whether the SafeStat model did a better job than random selection of identifying motor carriers that pose high crash risks.

In addition, we tested changes to selected portions of the SafeStat model to see whether improvements could be made in the identification of high-risk motor carriers. In one analysis, we modified the calculation of the safety evaluation area values and compared the number of high-risk motor carriers identified with the number identified by the unmodified safety evaluation areas. For example, we included carriers with only one crash in the calculation of the accident safety evaluation area whereas the

[1]We obtained crash data for this period that were reported to FMCSA through June 2006. This allowed us to obtain data on late-reported crashes for the July 2004 through December 2005 period.

[2]We drew 10,000 simple random samples of 4,989 carriers (the number of carriers that SafeStat identified as being at highest risk for crashes when we recalculated it) from the list of all carriers in the MCMIS master file used by SafeStat on June 25, 2004, and for each sample we calculated how many crashes the selected carriers reported to MCMIS between June 26, 2004, and December 25, 2005.

unmodified SafeStat model includes only carriers with two or more crashes. We also investigated the effect of removing the time and severity weights from the indexes used to construct the accident, driver, and vehicle safety evaluation areas. We then compared the result of using the modified and unmodified safety evaluation area values to determine if this modification improved the model's ability to identify future crash risks.

To assess the extent to which the timeliness, completeness, and accuracy of MCMIS and state-reported crash data affect SafeStat's performance, we carried out a series of analyses with the MCMIS crash master file and MCMIS census file, as well as surveying the literature to assess findings on MCMIS data quality from other studies. To assess the effect of timeliness, we first measured how many days on average it was taking each state to report crashes to FMCSA by year for calendar years 2000 through 2005. We also recalculated SafeStat scores from the model's June 25, 2004, run to include crashes that had occurred more than 90 days before that date but had not been reported to FMCSA by that date. We compared the number and rankings of carriers from the original SafeStat results with those obtained by adding in data for the late-reported crashes. In addition, we reviewed the University of Michigan Transportation Research Institute's studies of state crash reporting to MCMIS to identify the impact of late reporting in individual states on MCMIS data quality.

To assess the effect of completeness, we attempted to match all crash records in the MCMIS crash file for crashes occurring from December 2001 through June 2004 to the list of motor carriers in the MCMIS census file. In addition, we reviewed the University of Michigan Transportation Research Institute's studies of state crash reporting to MCMIS to identify the impact of incomplete crash reporting in individual states on MCMIS data quality.

To assess the effect of accuracy, we reviewed a report by the Office of Inspector General that tested the accuracy of electronic data by comparing records selected in the sample with source paper documents. In addition, we reviewed the University of Michigan Transportation Research Institute's studies of state crash reporting to MCMIS to identify the impact of incorrectly reported crashes in individual states on MCMIS data quality.

While the limitations in the data adversely affect the ability of any method to identify carriers that pose high crash risks, we determined that the data were of sufficient quality for our use, which was to assess how the application of regression models might improve the ability to identify high-risk carriers over the current approach—not to determine absolute measures of crash risk. Our reasoning is based on the fact that we used the same data set to compare the results of the SafeStat model and the regression models. Limitations in the data would apply equally to both results. Methods to identify carriers that pose high crash risk will perform more efficiently once the known problems with the quality of state-reported crash data are addressed.

To understand what other researchers have found about how well SafeStat identifies motor carriers that pose high crash risks, we identified studies through a general literature review and by asking stakeholders and study authors to identify high-quality studies. Studies included in our review were (1) the 2004 study of SafeStat done by Oak Ridge National Laboratory, (2) the SafeStat effectiveness studies done by the Department of Transportation Office of Inspector General and Volpe Institute, (3) the University of Michigan Transportation Research Institute's studies of state crash reporting to FMCSA, and (4) the 2006 Department of Transportation Office of Inspector General's audit of data for new entrant carriers.[3] We assessed the methodology used in each study and identified which findings are supported by rigorous analysis. We accomplished this by relying on information presented in the studies and, where possible, by discussing the studies with the authors. When the studies' methodologies and analyses appeared reasonable, we used those findings in our analysis of SafeStat. We discussed with FMCSA and industry and safety stakeholders the SafeStat methodology issues and data quality issues raised by these studies. We also discussed the aptness of the respective methodological approaches with FMCSA. Finally, we reviewed FMCSA documentation on how SafeStat is constructed and assessments of SafeStat conducted by FMCSA.

[3]Campbell, Schmoyer, and Hwang, *Review of The Motor Carrier Safety Status Measurement System (SAFESTAT)*, 2004; U.S. DOT Office of Inspector General, *Improvements Needed In the Motor Carrier Safety Status Measurement System*, 2004; Madsen and Wright, U.S. DOT-Volpe National Transportation Systems Center, *An Effectiveness Analysis of SafeStat*, 1998; John A. Volpe National Transportation Systems Center, *SafeStat Effectiveness Study Update*, 2004. University of Michigan Transportation Research Institute MCMIS State Reports; U.S. DOT Office of Inspector General, *Significant Improvements in Motor Carrier Safety Program Since 1999 Act But Loopholes For Repeat Violators Need Closing*, 2006.

Appendix III: Additional Results from Our Statistical Analyses of the SafeStat Model

This appendix contains technical descriptions and other information related to our statistical analyses.

Overview of Regression Analyses

To study how well statistical methods identify carriers that pose high crash risks, we carried out a series of regression analyses. The safety evaluation area values for the accident, driver, vehicle, and safety management areas served as the independent variables to predict crash risks.[1] We used the state-reported crash data in MCMIS for crashes that occurred during the 30 months preceding June 25, 2004, as the dependent variable in each model. We used the results of the SafeStat model run from June 25, 2004, to benchmark the performance of the regression models with the crash records for the identified high-risk carriers over the succeeding 18 months.

We matched the state-reported crashes that occurred from December 26, 2001, through June 25, 2004, to the carriers listed in SafeStat.[2] We checked our match of crashes for carriers with those carriers used by FMCSA in June 2004 and found that the reported numbers had changed for about 10,700 carriers in the intervening 2 years. We found this difference even though we used only crashes that occurred from December 26, 2001, through June 25, 2004, and were reported to FMCSA before June 25, 2004. Because of this difference in matched crashes, we recalculated the accident safety evaluation area using our match of the crashes. This is discussed later in more detail.

Using our recalculation of the accident safety evaluation area values and the original driver, vehicle, and safety management safety evaluation area values for the carriers, we fit a Poisson regression model and a negative binomial regression model to the crash counts. Both of these models are statistically appropriate for use when modeling counts that are positive and integer valued. The two models differ in their assumptions about the mean and variance. Whereas the Poisson model assumes that the mean and the variance are equal, the negative binomial model assumes the mean is not equal to the variance. The crash data in MCMIS fit the

[1] In addition to the safety evaluation area scores, we included indicator variables to flag any missing safety evaluation area scores.

[2] We used the carrier's Department of Transportation number recorded in the crash record to match to the carrier's Department of Transportation number listed in the SafeStat summary table.

assumptions of the negative binomial distribution better than those of the Poisson.[3]

We also tried to estimate zero-inflated Poisson and zero-inflated negative binomial models with the SafeStat data. These models are appropriate when the count values include many zeros, as is the case with the values in this data set (because many carriers do not have crash records). However, we could not estimate the parameters for these models with the MCMIS data. We also considered using logistic regression to model the carrier's odds of experiencing a crash. However, the parameter estimates of the four safety evaluation area values could not be estimated, so we did not use the results of this model.[4]

Finally, we used the results from the negative binomial model to assess the expected carrier crash counts using the empirical Bayes estimate. In safety applications, the empirical Bayes method[5] is used to increase the precision of estimates and correct for the regression-to-mean bias.[6] In this application, the empirical Bayes method calculates a weighted average of the rate of crashes for a carrier from the prior 30 months with the predicted mean number of crashes from the negative binomial regression. This method optimizes the identification of carriers with the highest number of future crashes. This optimization of total crashes, however, resulted in the identification of primarily the largest companies. The crash rate (crashes per 1,000 vehicles per 18 months) was not as high for this group as for the carriers placed by the SafeStat model in its A and B categories.

[3]We checked this by estimating the mean and variance of the crashes for the population of all carriers and determined that they were significantly different.

[4]The coefficients in the model could not be reliably estimated (the maximum likelihood of the model did not converge).

[5]Hauer, Harwood, Council, and Griffith, *Estimating Safety by the Empirical Bayes Method: A Tutorial*, 2001.

[6]In the context of crashes, we wish to "treat" the most dangerous companies with a compliance review to make them safer. But, crashes are distributed with a fair degree of randomness. A company selected for a compliance review may have suffered an atypical random grouping of accidents in the preceding months. With or without a compliance review, it is likely that the random grouping will not exist next year, and the crash figures will improve. Statistical methods seek to control for this regression-to-mean bias in order to better identify the effect of a compliance review on a company's safety.

Technical Explanation of the Negative Binomial Regression Model

This section provides the technical details for the negative binomial regression model fit to the SafeStat data. This section also explains how we handled incomplete safety evaluation area data for carriers in the regression model analyses.

The basic negative binomial probability distribution function for count data is expressed as

$$f(y_i \mid x_i) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma(y_i + 1)\Gamma\left(\frac{1}{k}\right)} \frac{(k\mu)^{y_i}}{(1 + k\mu)^{y_i + \frac{1}{k}}}$$

for $y_t = 0,1,2,3,\ldots$ The term k represents the dispersion parameter, $k > 0$. It is not assumed to equal one, as in the Poisson distribution. The y_i represents the crash count for the i^{th} motor carrier, and the x_i represents the observed safety evaluation areas. To formulate the negative binomial regression model and control for differences in exposure to events among the carriers, we can express the functional relationship between the safety evaluation areas and the mean number of crashes as

$$\mu_i = \exp(x_i^T \beta)$$

With complete data for a motor carrier, where none of the safety evaluation area values are equal to missing, the regression model of interest is as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 ACC_i + \beta_2 DRV_i + \beta_3 VEH_i + \beta_4 SMgt_i$$

This equation models the log of the expected mean number of crashes for each motor carrier using the four safety evaluation area values, but most commercial motor companies listed in MCMIS do not have values for all four safety evaluation areas.[7] To account for this, it is necessary to define four indicator variables. Let

$$I_{ACC} = \begin{cases} 1 \text{ if the accident safety evaluation area is present} \\ 0 \text{ otherwise} \end{cases};$$

$$I_{DRV} = \begin{cases} 1 \text{ if the driver safety evaluation area is present} \\ 0 \text{ otherwise} \end{cases};$$

$$I_{VEH} = \begin{cases} 1 \text{ if the vehicle safety evaluation area is present} \\ 0 \text{ otherwise} \end{cases};$$

and $$I_{SMgt} = \begin{cases} 1 \text{ if the safety management safety evaluation area is present} \\ 0 \text{ otherwise} \end{cases}.$$

The indicator variables will be used as main effects in the negative binomial regression model to indicate cases for which information is available. The effect of the safety evaluation area will be measured by the interaction of the indicator function with the safety evaluation area value. This gives us the following model specification:

$$\log(\mu_i) = \beta_0 + \beta_1 I_{ACC} + \beta_2 (I_{ACC} \times ACC_i) + \beta_3 I_{DRV} + \beta_4 (I_{DRV} \times DRV_i) + \beta_5 I_{VEH} + \beta_6 (I_{VEH} \times VEH_i) + \beta_7 I_{SMgt} + \beta_8 (I_{SMgt} \times SMgt_i)$$

With this parameterization, the estimate for the mean rate of crashes for a carrier with no safety evaluation area information is $\exp(\beta_0)$. For a carrier with information for just the accident safety evaluation area, the estimate for the mean number of crashes is $\exp(\beta_0 + \beta_1 + \beta_2 ACC_i)$. Note that the effect for each safety evaluation area will include a coefficient times the safety evaluation area value for the carrier plus an offset to the

[7]A carrier has to have two or more reported crashes in the past 30 months to receive an accident safety evaluation area value. A carrier has to have three or more roadside inspections to receive a driver or vehicle safety evaluation area value. A driver has to have had a compliance review in the past 18 months to receive a safety management safety evaluation area value. There are other ways a carrier can receive a value for one of these four safety evaluation areas, refer to the description of each one provided in the Background.

intercept for the indicator term (the coefficient for the indicator function).

We used a similar parameterization to formulate the Poisson regression model.

Evaluation of Regression Models' Performance

We estimated regression models using the same data FMCSA used in its application of the SafeStat model on June 25, 2004, with one exception for the accident safety evaluation area. For that area, we used our own match of crashes to carriers for December 26, 2001, through June 25, 2004. The MCMIS data we received in June 2006 produced different totals in the match of crashes to carriers for about 10,700 carriers. MCMIS data change over time because crash data are added, deleted, or changed as more information about these crashes is obtained. The discrepancies in matching arose even though we used the identical time interval and counted crashes only when the record indicated they had been reported to FMCSA before June 25, 2004. Because of these discrepancies, it was necessary to calculate the accident safety evaluation area values using our match of crashes and then recalculate the SafeStat carrier scores for June 25, 2004, using our accident safety evaluation area values and the original driver, vehicle, and safety management safety evaluation area values.[8] We used our accident safety evaluation area values and the original driver, vehicle, and safety management safety evaluation area values in the regression model analysis.

Using the revised accident safety evaluation area values and FMCSA's original driver, vehicle, and safety management safety evaluation area values, the SafeStat model identified 4,989 carriers that pose high crash risks. For each regression model, we input the safety evaluation area data for the carriers in our analysis data set and used the regression model to calculate the predicted mean number of crashes. We then sorted the predicted scores and selected the 4,989 carriers with the worst predicted values as the set of high-risk carriers identified by the regression model. Next, we used MCMIS to determine the crash history of these 4,989 carriers between June 26, 2004, and December 25, 2005, and compared

[8]Our calculation of the accident safety evaluation area differed slightly from that used by FMCSA. We did not add 1 to the severity weights for crashes with an associated hazardous materials release due to the rarity of this event.

the aggregate crash history with the aggregate crash history of the carriers identified by the SafeStat model during the same period of time.

The regression models do not categorize carriers by letter; the regression models produce a predicted crash risk for each carrier. The regression models make use of the safety evaluation area values, but they differ from the SafeStat model in this respect.

The results show that a negative binomial regression model estimated with the safety evaluation area values outperforms the current SafeStat model in terms of predicting future crashes and the future crash rate among identified carriers that pose high crash risks. (See table 3.) That is, our negative binomial and Poisson models show 111 and 109 crashes per 1,000 vehicles per 18 months, respectively, compared with the 102 crashes per 1,000 vehicles per 18 months estimated by the current SafeStat model. The Poisson model is not as appropriate since the crash counts for carriers have variability that is significantly different from the mean number of crashes.[9] The empirical Bayes method optimizes the selection of future crashes; however, it does so by selecting the largest carriers. The largest carriers have a lower crash rate per 1,000 vehicles per 18 months than the carriers that pose high crash risks identified by the SafeStat model or by the negative binomial regression model. Since the primary use of SafeStat is to identify and prioritize carriers for FMCSA and state compliance reviews, the empirical Bayes method did not identify carriers with the highest safety risk.

[9]The equality of the variability in the number of crashes to the average number of crashes is an assumption of the Poisson regression model. This assumption does not hold for the MCMIS data we analyzed.

Table 3: Results for SafeStat Model and Regression Models

Method	Crash rate	Number of crashes in 18 months	Number of vehicles
SafeStat category A & B	102	10,076	98,619
Negative binomial	111	19,580	175,820
Poisson	109	21,532	198,396
Empirical Bayes	59	56,705	965,070

Source: GAO analysis of FMCSA data.

Note: As discussed in the text, the zero inflated Poisson, the zero inflated negative binomial, and the logistic regression approaches did not provide useful results.

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548