HOURS OF SERVICE REGULATORY IMPACT ANALYSIS:

PEER REVIEW RESULTS AND FMCSA RESPONSES

MAY 2008

Summary

FMCSA submitted the December 2007 Hours of Service (HOS) Regulatory Impact Analysis (RIA) to a peer review by three experts in the areas of motor carrier safety, driver behavior, and sleep and driver performance. The reviewers were Linda Ng Boyle, PhD, Assistant Professor, Department of Mechanical and Industrial Engineering, University of Iowa; Ron Knipling, PhD, Senior Research Scientist and Senior Transportation Fellow, Virginia Technical Institute Transportation Institute; and Dr. Greg Belenky, M.D., Research Professor and Director, Sleep Performance Research Center, Washington State University. Due to the time constraints of this review, the review team concentrated on the regression modeling technique used in the Time on Task (TOT) analysis; the use of the FAST/SAFTE model for assessing the relationships between driver fatigue, sleep, and safety performance; and crash and fatigue data and studies used or cited in the analysis.

The peer review was generally favorable to this latest HOS RIA. The review team offered some suggestions on the logistic regression model used in the analysis, but none that would lead the Agency to significantly alter its specification. The team did validate this particular type of model's applicability to the TOT analysis. The reviewers judged the use of the SAFTE/FAST to be generally "sound", but suggested that modeling of the effect of irregular hours and sleeper berth use should be improved. They also highlighted idiosyncrasies of the data the Agency used in its analysis. The review team made an important suggestion that time awake is a better predictor of fatigue than TOT, and in that

light, states that the reduction in overall on-duty times implemented under this rule, insofar as it has decreased overall time awake, is much more significant than the increase in drive time.

What follows are point-by-point responses to the peer reviewers comments.

Responses to Peer Review Comments

Comments on Validity of Model

The reviewers provided several comments on the linear Time on Task (TOT) model used in the 2005 HOS RIA. However, because the linear model used on that analysis is not being used in the current RIA, FMCSA will not address these comments. The original linear model was determined to be flawed because it allowed for the predicted probability of a fatigue-related crash to be greater than one. It was subsequently determined that a logistic model used is more appropriate for modeling TOT effects where the dependent variable is a probability.

A reviewer claimed that the descriptions of the modeling technique are incorrect, and state that the logistic model is set up to predict the probability of an event given the absence of the non event, and should be specified as:

$$\log P\left\{\frac{\text{(fatigue related crash)}}{\text{(non fatigue related crash)}}\right\} = \beta_0 + \beta_i X_i + \varepsilon$$

Response

The above statements and notation are incorrect. The proper logistic model would have the form:

$$\log\left\{\frac{P(\text{fatigue related crash})}{P(\text{non fatigue related crash})}\right\} = \beta_0 + \beta_i X_i$$

In the analysis, logit(p) is defined as logit(p)=log{p/(1-p)}. Therefore, the model presented, $\text{logit}\{P(\text{fatigue related crash})\} = \beta_0 + \beta_i X_i + \varepsilon$, is equivalent to a standard logistic model. The RIA text has been edited to clarify this point.

Comments on Alternative Model Specifications

A reviewer commented on modeling the effect of time of day on the probability of a fatigue related crash. According to alternative specifications of the logistic model, the probability of a fatigue-related crash increases during driving between 22:00 and 09:00, and the probability of a fatigue-related crash decreases with driving from 10:00 to 21:00. This reviewer asks if these results are significant and how they compare with actual crash data. The commenter also suggests that the model should include interaction terms between, for example, hours of driving and time of day, and states a belief that the coefficients on these terms would be significant.

Response

The phenomenon of fatigue-related crashes increasing late at night and into the early morning hours, and decreasing during the day, accords well with theory, both because circadian rhythms normally reach a low in the early morning, and because many drivers will have been awake for many hours if they are driving after midnight. Thus, there are no reasons to suspect that this approach is incorrect. Statistically, the Wald p-value calculations show that time of day is a statistically significant variable, and we could test whether the difference between the probabilities of fatigue-related crashes at any given two times of the day is statistically significant, all else being equal. We did not investigate more complex models involving interactions, for example, between time of day and hours of driving. The additional predictor variables did not change the average effects of hours of driving estimated in the TOT model. The suggestion of including interaction terms will be considered for fitting additional models.

Comments on Application and Interpretation of Model

One reviewer had some general comments about how the logistic model was applied, interpreted, or explained. She commented that the regression model cannot be extrapolated to values that are not in their predictor region, and suggests that a time series analysis would be appropriate for predicting future observations. Because no time series analysis was conducted, the reviewers suggest that the authors remove this claims about and results of extrapolations. This reviewer also questioned whether the confidence intervals around the predicted values had been calculated properly, because the analysis did not give enough detail on these calculations, and points out that the confidence interval for the probability estimate 1.9% is not reported. She also pointed out an error in a statement made on page 45 of chapter 5 claiming that "the statistical significance of the [11th hour] results are(sic) rather low, which is presumably causing the very high variance surrounding the estimated 11th hour crash risk." Low statistical significance does not cause high variability.

Response

The analysis does not claim to predict future values, that is, to forecast values. Nevertheless, all fitted values are within the TOT hours in the data set, which extend from 1 to 36. The reference to extrapolation was from the Circuit Court decision:

> "The agency, for example, could have extrapolated the time-on-task effects of driving longer hours using crash-risk data derived from drivers who drove for shorter periods of time."

The text will be revised to reduce the chance of misunderstanding. The point is that all of the available data are used to estimate the probabilities at hours of driving even where the observed number of crashes was zero or small. The available TIFA crash data is heavily concentrated among low hours of driving.

The confidence intervals are not symmetric and were calculated correctly in a fashion analogous to what the reviewer has described, but were converted to probabilities. Specifically, the log-odds predicted by the model is calculated as pi, and converted to a probability by the equation 1/(1+exp(-pi)). The text will be modified as appropriate.

The probability estimate of 1.9% was presented as an example of the application of a model 2, a variant of the original model. The confidence intervals were calculated only for the primary logistic model.

The erroneous text on page 45 will be revised. The sentence was intended to mean that the small number of data points led to both high uncertainty about the 11th hour crash risk and the low significance.

Comments on Data

A member of the review team remarked that explanations of how data were constructed or adjusted were confusing or incomplete. The analysis does not clearly explain why the TOT fatigue function was divided by 2.92%, or how it does not affect the results of the analysis (and if it has no effect, why it was even applied). Although the equations are shown to be valid, the commenter states that is more important to clearly explain where did each variable came from, for instance, how the fatigue percentages at the 10_{th} and 11_{th} hours were measured, where the vector of risk estimates and the weights for the risks were derived from.

Response

The adjustment in the fatigue results was needed to correct for the fact that original analysis used an incorrect TOT function and did not apply it consistently. The measured fatigue percentages are from the logistic model described in Appendix V, and the risk estimates and weights are from the fatigue and simulation models described in Chapter 5. These sources are mentioned in the text, but will be clarified.

Comments on Alternatives to TOT Models

A reviewer strongly advocated modeling time awake instead of TOT, and cited several studies that followed this approach. He stated that TOT effects found in various research and crash analysis studies may largely be artifacts of time awake. This distinction bears directly on the HOS rules because sleep schedules are determined by work schedules. This commenter also noted that within the normal 16-hour wakefulness period, the actual task performed does not strongly affect alertness unless the task is extremely strenuous or demanding. He adds that in light of time awake effects, the positive impact of the decrease from 15 to 14 work hours is more significant than the extension of driving hours from 10 to 11.

This reviewer also notes that other data sets may provide very different pictures in relation to TOT effects on crash risk. He presents a comparison of exposure (hours of driving) percentages from the RIA to hours-of-driving in several crash categories from the Large Truck Crash Causation Study (LTCCS) conducted by FMCSA. His analysis finds no TOT effect for the three categories of crash involvements examined: all crash involvements, all involvements with the Critical Reason assigned to the truck, and single-vehicle truck crash involvements. However, he cautions this analysis is not definitive because the crash and exposure data come from different and unrelated sources, but it does suggest that the TOT-to-crash risk relationship is not robust or validated.

Response

Some TOT effects might be due to time awake rather than hours of driving, but the data available at the time the analysis was conducted did not allow us to separate them. The assumption that all of the effects of TOT observed in the data were due to driving hours was considered to be conservative. Limiting working hours to 14 the 2003 HOS rules was superior to the previous regime.

Confirmation of the LTCCS analysis this reviewer presented is underway. Again, the use of an independent TOT factor, based on the TIFA data, can be considered conservative.

General Comments on SAFTE/FAST Model

A reviewer states that current research has validated the use of the SAFTE/FAST model (Hursh et al., 2004; Hursh et al., 2006; Dean et al., 2007). The Agency will reference this research these papers.

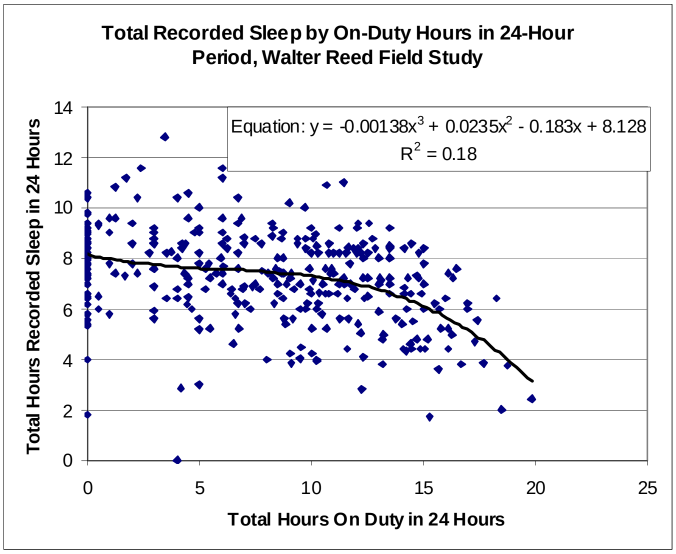
Comments on Sleep Estimate

A reviewer questioned the source for the estimates of sleep times of 6.57 hours in 10 hours off duty, and stated that it seemed low.

Response

The estimate of 6.57 hours of sleep for 10 hours off duty was derived for the 2005 HOS RIA. The amount of sleep a driver is expected to get is based on past work history, and is estimated from a function that decreases as the cumulative amount of on-duty time in the previous 24 hour period increases. The basic function is identical to the one used in the 2003 RIA.

Section 8.4 of the 2003 HOS RIA explains in greater detail how the amount of sleep was estimated. A cubic regression equation (third-order polynomial) was fitted to actigraph data from the Walter Reed field study to predict sleep in a 24-hour period based on time on duty in 24 hours. This model was used to capture the apparent non-linearity of the data; the amount of sleep falls at an increasing rate as the number of on-duty hours increases. The following chart and table show the original data and the results from the modeling.



Hours On Duty	Predicted Hours of Sleep
0	8.13
6	7.57
8	7.45
10	7.25
12	6.91
14	6.35
16	5.52
18	4.34
20	2.75

The hours of sleep predicted by this function are consistent with experimental and real-world observations. For example, the Driver Fatigue and Alertness Study (Wylie, Shultz, Miller, Mitler, Mackie, 1996), also found the truck drivers got relatively little sleep. A sample of drivers averaging 10.70 hours off between drives reported 5.78 hours in bed in their principal sleep periods (Table 4-3 on p. 4-14and Table 4-5 on p. 4-19. The study noted on p. 4-20 that "Even after accounting for study protocol related demands, drivers, for undetermined reasons, used too little of their available free time for sleep." FMCSA also reported a low number of hours of sleep (6.28 hours) under its new rules.

Comments on Application of SAFTE/FAST Model

A reviewer commented on the application of the SAFTE/FAST model to the issue of length of restart provision. This commenter states that the argument for the 34 hour restart is not unreasonable considering that the model did not find appreciable differences between the two restart conditions, a 34-hour and a 58-hour restart, it tested. However, more conditions should have been modeled; for example; the report should have modeled a schedule in which during the 6 days the on duty period was nocturnal and the off duty time was diurnal, i.e., schedules in which the sleep opportunity was placed at a time of day less conducive to getting sleep. Because the modeling effort was too limited to produce definitive results, this reviewer suggests that it be redone systematically with a larger number of possible schedules.

This reviewer also commented on the use of the SAFTE/FAST model for assessing the effects of irregular vs. regular schedules on driver performance. A regular schedule is one where the driver works and rests at the same general time of day over consecutive days. The report presented two SAFTE/FAST graphical plots, one showing an optimal regular schedule with 8 hours of sleep from 23:00-07:00 and on duty from 08:00-19:00, and the other showing an irregular schedule with the sleep intervals generally shorter and occurring at different times on successive days. Not only was sleep shorter in the irregular condition but SAFTE/FAST reduces the recuperative value of sleep taken at a non-circadian friendly time of day, and unsurprisingly, performance is degraded with the irregular schedule. This commenter notes that one can imagine a suboptimal regular schedule with sleep from 11:00-19:00; in this case recuperative value of the regular diurnal sleep would be reduced and SAFTE/FAST could quite conceivably find an irregular schedule supporting better overall performance (especially if total sleep times remained the same in both conditions). The reviewer suggests that the modeling of regular versus irregular schedules should be redone or removed because it is poorly constructed and involves too few schedules to be definitive.

Response

The analysis of the restart provision did indeed directly compare the fatigue and safety effects of shorter and longer restart breaks, but the cases cited were not the only basis for the conclusions about the benefits of longer restart breaks. Rather, they were presented as illustrations of the effects of longer restart breaks, and as particularly "pure" cases for comparison. The report's conclusions were based on modeling of fatigue levels under a large number of scenarios, generated by the operational simulation model. Furthermore, the SAFTE/FAST model assumes that drivers on a night driving/day sleeping schedule will adjust to it so that its predictions should be the same whether a continuous day or continuous night schedule is modeled.

Likewise, the two work schedule scenarios were presented for illustrative purposes; the actual estimates were based on a large number of scenarios that were generated by a model intended to mirror real-world conditions. As stated, SAFTE/FAST is designed to allow a driver's circadian rhythms to shift so as to be "friendly" for driving and sleeping at times of day and night that correspond to that driver's schedule. Thus, the scenario that the reviewer proposes should not be different from the one that was presented. Recent work done by Belenky and Hursh showing that a split schedule is a more robust way to ensure that drivers are getting sleep at optimal times (even though their sleep is interrupted) did not model a long enough period of time to allow a night driver to adapt to a day sleeping schedule.

Comments on Analysis of Sleeper Berth Provisions

A reviewer commented that the modeling of sleeper berth use is poorly constructed and unsystematic and either should be redone properly or should be removed from the 2007 RIA. This commenter stated that the analysis failed to allocate all the available off-duty time, and refers to an analysis of 14 hour on-duty schedules which found that a split sleeper-berth use was optimal under diurnal sleep conditions, and that this type of sleeper-berth use was on average better than consolidate sleep.

Response

The case that was modeled (in which the drivers worked 10 rather than 14 hours per day) was more realistic than the 14-hour workday case recommended by the reviewer. Information from industry experts indicates that team drivers tend to work less than the maximum allowed amount in a given day because there are long periods in which the truck is either stopped or the other driver is at the wheel. As stated, the SAFTE/FAST model assumes that drivers adapt to a shifted schedule, so that each one is likely to perform better with consolidated sleep than with split sleep, assuming that periods between trips are short enough to prevent the night-adapted driver from losing all of his or her adaptation.

Comments on Individual Differences in Fatigue Susceptibility

One reviewer referred to multiple studies that have indicated sharp individual differences in human susceptibility to drowsiness. This commenter acknowledges that, although HOS rules are not likely to incorporate individual differences in the foreseeable future, this factor should be considered because future technologies may permit rules to be tailored to individual ability to sustain alertness, and the reality of large individual differences belies the fact that there is no direct relation between conventional HOS parameters and driver alertness and performance.

Response

None of the options considered allow for individual differences and so this point would not change the analysis.

Comments on Crash Data

A reviewer cautioned the Agency to avoid mixing statistics on driver fatigue for crashes of different severity levels. Crash severity increases with increases in driver impairment. To illustrate this point, the reviewer produced a table from Fatality Analysis Reporting System (FARS) data showing significantly higher percentages of fatigue related crashes among fatal crashes as compared to all crashes in general.

This reviewer also comments that the 7.25% baseline fatigue-related crash percentage is hard to rectify with statistics he presented or more recent FARS and GES fatigue-related crash percentages. For example, the 2003 and 2004 FARS percentages for fatigue-related fatal crashes were 1.7% and 1.5%, and the GES percentages were even lower. The commenter suggests that the 7.25% figure might be derived by eliminating unknowns and crashes where other drivers were primarily at-fault, although the figure is purported to be for all truck-involved crashes. He further points out that Knipling and Shelton (1999) estimate that fatigue-related crashes would account for 2.8% to 6.1% of fatal crashes and 0.5% to 1.1% of all crashes if more thorough investigations crashes were to occur.

This reviewer states that there is no systematic rationale for the use of an increment of 0.89% for inattention because, although FARS certainly greatly understates the role of inattention in crashes, there is no apparent basis for asserting that 20% of them are related to fatigue; naturalistic driving studies show much more driver inattention but also show that inattentive drivers are usually awake and looking away from the forward road scene during inattention episodes.

The reviewer also expressed a concern about "circularity" in the crash data. In crash investigation studies (including FARS, TIFA, GES, NTSB studies, and the LTCCS), the investigator's "fatigue" determination is partially based on the same schedule variables that are later analyzed as factors causing fatigue. The reviewer acknowledged that there is no ready alternative to these crash datasets, but suggested that these difficulties should be discussed, and proposes that FMCSA consider combining more objective and rigid crash determinations with actual exposure data, perhaps generated through naturalistic driving experiments

Response

The Agency is aware of the limitations in the data mentioned by this reviewer. FARS data are not ideal for illuminating the effects of fatigue on crashes, given that fatal crashes will be examined more thoroughly by investigators. Most sources discounted the very low rates of fatigue-related crashes seen in FARS and noted by this reviewer. A careful review of the FARS data for the 2003 HOS RIA found a much higher rate after cases in which fatigue was not considered were removed, and the LTCCS (also a more careful assessment) found a number comparable to the baseline fatigue-related percentage used in the analysis.

As reported in the 2003 HOS RIA, the 20% fatigue-related value for the inattention crashes increment was based on a study conducted at Virginia Tech: Hanowski, R., Wierwille, W., Gellatly, A., Early, N., and Dingus, T. *Impact of Local/Short Haul Operations on Driver Fatigue: Final Report,* FMCSA, FMCSA No. DOT-MC-00-203, NTIS No. PB2001-101416INZ, Washington, D.C., Sept. 2000.

The Agency shares the general concerns about crash data and studies, but believes that it was appropriate to use the TIFA-based analysis and consider its use to be conservative in that it might well overstate the increased risks of long driving hours.

Comments on Other Studies

The reviewer also commented on the preliminary Jovanis study cited in the RIA. He noted that the LTCCS shows that 3.9% of truck crashes had asleep-at-the-wheel as the Critical Reason, implying that only about 9 of the Jovanis crashes would have been primarily sleep-related, not enough for a valid trend analysis. Given the heavy preponderance of non-fatigue crashes in the dataset, it is likely that observed trends reflect primarily non-fatigue-related factors. In regard to the 11th hour, the reviewer states less than 2 percent—only 4 out of 231 crashes—of all the data presented by Jovanis are relevant.

Response

Regarding the comments on the preliminary Jovanis study, this study was cited, but not used, in the 2005 or 2007 RIAs. The Agency, however, shares the reviewer's concerns about this study. Regardless, asleep-at-the-wheel recorded in the LTCCS and fatigue are related but not identical concepts. The Agency cannot comment on the fraction of relevant data because it does not know how this estimate was derived.