Proposal for Non-Randomized Study of the Effectiveness of FAA Proposed Regulations Regarding Cockpit Crew Fatigue

August 16, 2011

Donald B. Rubin, PhD
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, Massachusetts 02138

Telephone number: 617-495-5498
Email address: rubin@stat.harvard.edu

This document describes the methodological steps that are needed to conduct a study, using existing observational data, to estimate the causal effects, both intended and unintended, of the FAA's proposed regulations on flight crewmember duty and rest requirements. (A summary of the FAA proposal, from September 2010, is attached as an Appendix to this Proposal.)

The Use of Randomized Experiments – in General

In outlining those steps, it is important to understand that the challenge facing FAA is not unique. There are numerous comparable situations involving federal regulatory agencies who must estimate such causal effects, many involving similar "interventions" that propose to change such things as the standard of medical care or to implement new job-training programs. Examples commonly occur within the context of the U.S. Food and Drug Administration's (FDA) approval process for new pharmaceuticals for treatment of diseases or devices for the treatment of medical conditions, or within the context of considering implementing new educational policies. Specific earlier examples include the recommendations for how to treat certain types of breast cancer: should mastectomy be the standard of care, or should breast conservation be preferred in some circumstances? Or, is a new cholesterol-reducing drug safe in general use and effective for reducing heart disease?

With such questions, there will not be generally accepted, scientifically-based answers; they are absent for a variety of reasons. These include the presence of vaguely understood causal factors operating that may affect the primary outcomes and the unanticipated negative side-effects of new interventions. The literature discussing interventions with humans is replete with examples.

In general, it is widely accepted that the most valid and reliable evidence about causal effects is obtained from randomized experiments, in which some of the units being studied are randomly assigned to the new intervention, here called the active treatment group, and the others are assigned to the other one, here called the control group; for simplicity of description, we assume that the choice is between two conditions, the active treatment and the control, but in general there can be more than two treatment conditions, especially when searching for a combination of factors that appears to be most favorable.

Often, however, such randomized experiments can be conducted only in specialized populations of units, and as a result the information garnered from them needs to be supplemented with evidence on the causal effects of the proposed new intervention, relative to the control, from non-randomized or observational data. A specific example of such a situation -- involving the study of mastectomy versus breast conservation for the treatment of breast cancer -- is summarized in my (1997) description of a United States General Accounting Office (GAO) investigation into that question using both randomized and nonrandomized data.[1] In other common situations, there is no opportunity to investigate the effects of proposed interventions using randomized experiments. This inability to randomize the units under study may be due to ethical concerns, or to the long lead time needed to obtain answers from yet to be conducted randomized experiments, or to practical obstacles created by implementation issues such as the costs involved, financial or otherwise.

For instance, it might be inappropriate or unwise to have flight crews on scheduled passenger flights work prolonged shifts, or operate under otherwise arduous conditions, for the purposes of experimentation about flight and duty period limits. Consider what happened in 2006 when, according to press reports, FAA officials permitted the airline JetBlue to outfit a small number of pilots with devices to measure alertness and then to assign some crews to work longer shifts in the cockpit -- as many as 10 to 11 hours a day. The study was designed to assess whether pilots could safely fly far longer without exhibiting ill effects from fatigue. After the study was disclosed, the airline and FAA staff were roundly criticized, with an FAA official stating, "We don't allow experiments with passengers on board, period."[2] If this study had instead randomly selected crews to work shorter versus normal shifts, the havoc created with airline schedules might have made the experiment impractical. Similarly, because of the costs incurred by air carriers in shortening the work day, subjecting some airlines, but not others, to more restrictive rules in a randomized study might be unacceptable for competitive reasons.

Although there are many situations where using randomized experiments to estimate causal effects of proposed interventions may not be possible, it is imperative that the answers from studies of *non*-randomized data nevertheless approximate, as closely as possible, the answers that *would* have been obtained from the corresponding randomized experiment. In fact, this was the unanimous conclusion of the opening

---

[1] See Rubin, D., *Estimating Causal Effects from Large Data Sets Using Propensity Scores,* Annals of Internal Medicine 127:(Supplement 2) 757-763 (1997); *Breast Conservation Versus Mastectomy: Patient Survival in Day-to-Day Medical Practice and in Randomized Studies,* U.S. General Accounting Office, Report No. GAO/PEMD-95-9 (1994).

[2] See A. Pasztor and S. Carey, *Pilot-Fatigue Test Lands JetBlue In Hot Water: Airline Pushed FAA Limits On Cockpit Time but Failed To Tell Passengers on Planes,* The Wall Street Journal (Oct. 21, 2006), p. 1.

session (on July 31, 2011) of the 2011 Joint Statistical Meetings (JSM) in Miami, which was attended by representatives of Federal Statistical Agencies and their regulatory counterparts (including the FDA and the Centers for Disease Control), academia (Harvard University), and research organizations (National Institute of Statistical Sciences). And this is the situation that we find ourselves in when conducting an evaluation of the proposed FAA regulations. The general approach being advocated for causal inference is now widely known as "Rubin's Causal Model" (Holland, 1986) for work initiated in Rubin (1974, 1975, 1977, 1978), and continuing through Rubin (2008, 2010).

A Hypothetical Randomized Experiment of the FAA proposed Regulations

Even though we cannot conduct randomized experiments to evaluate the effects of the proposed FAA regulations, it greatly facilitates the identification of crucial ingredients of a non-randomized evaluation to consider, at least briefly, how a randomized evaluation would be conducted.

The first task is to identify the units of study that would be impacted by the new regulations: these would be individual flight segments. In a randomized experiment, a fraction, say half, of those segments would be randomly selected to be subject to the new regulations; the remainder would not be subject to the new regulations but only to the current regulations, the current "standard of care." In the aforementioned study of the treatment of breast cancer for a certain category of woman, the units were women with the disease, where the then-current standard of care was mastectomy, whereas the new intervention was breast conservation. The randomization there would ensure that the units subject to the new intervention and those subject to the control intervention had the same distributions of all pre-randomization characteristics, such as age, baseline blood pressure, education level, marital status, region of the country, etc. Baseline variables like these are commonly called "covariates" because they "co-vary" with the outcome variables, such as mortality, that are of primary interest in the study of the intervention.

In our setting, the units are flight segments, each one of which is currently operated under current regulations but would be operated under the proposed regulations were they in place. Thus, if we could conduct a randomized evaluation for the effect of the new regulations, we would toss coins to decide which of the segments would take place under the old regulations and which take place under the new regulations. We would then record all sorts of outcome measurements on each flight segment, including anything related to flight safety. To improve the precision of the comparisons between outcomes under the new and old regulations, and to allow for the detection of differing effects of the new versus old regulations under differing baseline conditions, we would also record extensive covariate data. These data are information about each flight segment that is unaffected by the assignment to be operated under the old versus new regulations, such as the type of aircraft, the experience of the crew, the weather conditions, the planned flight time for the segment, etc. Importantly, *before* the examination of any outcome data, descriptions of the planned analyses of that data under the two regulations would be provided and agreed on. This plan would ensure a fair assessment of the implementation of the new regulations in place of the old ones.

Even though a randomized experiment addressing the FAA question cannot be conducted, at least at present, thinking about how to design one thus provides important guidance on how to evaluate the specific proposed intervention. If it were possible to conduct a randomized experiment, ideally many factors defining possible interventions would be established first, and a factorial experiment (a way of evaluating the effects and possible interactions of several factors or independent variables), or better still, a fractional factorial experiment would be designed to discover the best combination of factors. *See* Box, G. E., Hunter, W.G., Hunter, J.S., and Hunter, W.G., *Statistics for Experimenters: Design, Innovation, and Discovery*, 2d Ed., Wiley, 2005, ISBN 0471718130

By using existing FAA and industry data, and with the anticipated support of FAA, OMB and industry and union representatives, the study of these non-randomized data described in this proposal could be designed to achieve answers that are fully objective and that approximate the answers that would have been achieved with a randomized experiment. We would take advantage of the array of specific techniques that have been developed over the decades to ensure the integrity of non-randomized studies, including propensity score methods,[3] principal stratification,[4] and various methods for

[3] The propensity score is the probability of a unit (e.g., person, classroom, school) being assigned to the active treatment in a study given a set of known covariates. Propensity scores are used to reduce selection bias due to the lack of random assignment in an observational study by creating similar groups based on these covariates. Paul Rosenbaum and I introduced the propensity score to provide an alternative method to help estimate treatment effects when treatment assignment is not formally randomized, but can be assumed to be unconfounded (Rubin, 1975). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. Biometrika, (1983) 70, pp. 41-55. (With P. Rosenbaum).

[4] Many scientific problems require that treatment comparisons be adjusted for post-treatment variables; however, the quantities being estimated by standard methods of adjusting for such variables are not causal effects in general, but only in special cases. To address this deficiency, I developed, with Constantine Frangakis, a general framework for comparing treatments adjusting for post-treatment variables that yields principal effects, causal effects within principal strata. Principal stratification with respect to a post-treatment variable is a cross-classification of subjects defined by the joint potential

matching, such as Mahalanobis metric matching, [5] and propensity score sub-classification. The ultimate goal in applying these methods is to estimate objectively what the effects of the proposed regulations would have had on safety had they been applied retroactively to flight segments that took place under existing rules. The purpose of the matching and sub-classification is to create, in the observational data, a group of treated units, *i.e.*, flight segments that would have been forbidden under the new regulations, and a "matching" group of control units, *i.e.*, flight segments that would not have been forbidden under the new regulations, where "matching" means having the same (or nearly the same) distribution of all relevant covariates.

Steps Required for An Objective Evaluation of the Proposed Regulations

Our proposed analysis would proceed in six discrete steps, with an optional seventh step:

1. **Initial discussions**. In order for the project to succeed optimally in practice, it is essential to have 'buy-in' from all stakeholders on the appropriate methods and metrics of the study. To accomplish this consensus, we would initiate discussions among the researchers, industry representatives, labor representatives, OMB, the FAA, and other entities to determine what outcomes would be measured and what variables, including baseline covariates, would need to be considered.

   For instance, covariates such as flight crew experience, age, training background (civilian versus military), and commuting practices might be considered. In terms of selecting outcomes for measurement, discussions might include how to address various definitions of safely completed flights; for example, a flight crew may opt to execute a missed approach or "go-around" procedure to abort a landing attempt before finally landing safely. A missed approach or go-around, in certain circumstances, might be indicative of crew performance issues. In other cases, a missed approach or go-around may indicate textbook decision

(continued...)

values of that post-treatment variable under each of the treatments being compared. "Principal Stratification in Causal Inference." *Biometrics,* (2002) 58, 1. pp. 21-29. (With C. Frangakis).

[5] In statistics, the Mahalanobis metric refers to a distance measure introduced by P. C. Mahalanobis in 1936. It is a method that incorporates correlations between variables. It is the same as Euclidean distance if the variables are uncorrelated, and it is also a scale-invariant method , i.e., not dependent on the scale of measurements.

making and execution on the part of the crew, whereas, in many other cases, a flight may be entirely uneventful. The initial discussions would address how the study should account for such variations and definitions.

It is important that all parties with knowledge and perspective to contribute on these subject areas be included in discussions on what outcome variables are to be measured, what covariates should be included, the manner in which various data sources will be used, and the analytical methods to be utilized. The inclusion of all parties will lead to a better-informed and more refined study and to a shared understanding among all parties that the sole purpose of the study is to obtain the most accurate and objective answer possible.

2. **Data Gathering.** Upon conclusion of initial discussions, we would begin gathering data from existing sources and identify gaps requiring *ab initio* collection efforts. It is our understanding the FAA itself already collects vast amounts of data through existing programs that could be used for the study being proposed. Moreover, we understand that a great deal of additional operational data, relevant to the study, is kept by air carriers.

   For instance, the internationally recognized Line Operations Safety Audit ("LOSA") appears to be an excellent source of data. LOSA is predicated on the notion that "understanding the human contribution to successes and failures in aviation can be better achieved by monitoring normal operations, rather than accidents and incidents." *See* Line Operations Safety Audit (LOSA), Doc 9803 AN/761 (1st Ed. 2002), International Civil Aviation Organization, at 1-2. LOSA consists of monitoring crew performance during all phases of normally conducted flights by an observer in the cockpit. Unlike other data sources, the generation of LOSA data is entirely independent of a specific outcome. It is not triggered by an accident or an incident. Thus, LOSA appears to provide a wealth of data to understand how flight crews perform in a spectrum of circumstances—including circumstances in which a flight crew member may experience fatigue but that does not result in an accident nor measurably reduce observable margins of safety, and also those circumstances in which a crew member experiences fatigue and consequently makes errors that might lead to an accident in certain circumstances. LOSA data certainly appear promising.

   Another promising source of FAA data consists of reports from the Aviation Safety Action Program ("ASAP"). The ASAP program collects voluntarily submitted reports of safety lapses from airline employees. Although ASAP reports are outcome triggered and possibly suffer from biases because of their voluntary nature, they nonetheless are a source of data about systemic hazards that could be used in a statistical analysis.

There are numerous other data sources such as the Flight Operational Quality Assurance ("FOQA") program, which provides data directly from the aircraft recorders and could capture instances of unsafe conditions, such as an unstable approach, which could indicate degraded pilot performance.

The FAA's Aviation Safety Information Analysis and Sharing ("ASIAS") System ingests the data sources discussed above, as well as information from 43 other databases. We propose to consider all available FAA and available industry data to determine what sources could be used in our analysis. We note that some data may be subject to confidentiality protections inuring to individuals or employee groups. At this stage, we are unable to ascertain what additional data collection might be required. The final decisions about the data sources to be used, and how they should be assembled, should involve subject matter experts, and should have buy-in approval from stakeholders before any analysis of outcome data for causal effects is initiated.

At the conclusion of this phase, the data should be separated into two parts, covariate data and outcome data. The outcome data would be safeguarded and securely hidden from study participants by a third party until the completion of all the analyses that are aimed at reconstructing the hypothetical randomized experiment with matching treatment and control flight segments with respect to the covariate data. Segregation of outcome data is necessary to ensure the objectivity of the resulting estimates of the causal effects of the proposed new regulations. This fact has been emphasized in several recent publications by me (Rubin 2008, 2010) and others (D'Agostino and D'Agostino, 2010, NEJM), and, as stated earlier, was unanimously agreed on in the opening session of the recently concluded 2011 JSM.

3. **Segment Classification as Treatment vs. Control.** The third step would be to classify the individual flight segments into 'control' and 'treatment' groups of flight segments. Flight segments falling into the 'control' group are those that would remain legally permissible even if the FAA's proposed regulations were applied. The treatment group segments are those flights that would become legally impermissible with the application of the FAA's proposed regulations. For instance, a crew at an all cargo airline can fly five consecutive nights under existing regulations. If the FAA proposal were adopted, they would be limited to three consecutive nights. Thus the flights operated on nights 1, 2 and 3 would

be in the control group assuming those segments satisfied the other parts of the new regulations, whereas the flights on nights 4 and 5, being prohibited, would thus be classified as being in the treatment group. This exercise would reflect the proposed regulations, as summarized in the included appendix, and would be implemented with the agreement of the stakeholders by third parties using approved software. Moreover, whenever any doubt arose as to the methods used, the stakeholder agreement would provide for a formal method of adjudication. This is the same process used in randomized medical experiments for adjudicating so-called adverse events, when the question is whether they might be related to medical interventions.

4. **Segment Matching.** Once flights have been classified as treatment or control segments, they must be matched on the covariates selected in Step 2, deemed probative of flight crew performance and flight safety in general. For example, flights could be matched according to: aircraft type, origin & destination airports, the ages of the crew, the season of the year, weather encountered, commuting practices of the particular crews, and airline-specific training and procedures. To the extent that exact matches on all attributes are not found, and this is extremely likely if there are more than a few covariates being considered, substantially similar flights can be paired according to pre-determined statistical methods. It is very likely that propensity score matching and sub-classification will be used, as well as other related techniques, some of which are summarized in the recent book by me, Rubin (2006) on *Matched Sampling for Causal Effects*. We can assemble teams of individuals with substantial experience doing such matching exercises for federal agencies, foreign regulatory agencies, and private companies. These techniques allow for the meaningful comparisons of treated and control flight segments to estimate the consequences of the proposed FAA regulations.

   At this point the created matched groups must be assessed for balance with respect to all covariates that are considered relevant. This assessment can be time consuming, because many such matched samples will probably have to be considered and assessed for balance. It is also imperative that all stakeholders provide their opinions on the adequacy of any final matched sample. Recall that all of this matching and assessing of balance is being conducted without any party (except for the holder of outcome data) having access to outcome data, and therefore without the benefit of answers about the estimated causal effects of the proposed regulations. The stakeholders' buy-in on the balancing of the groups -- that is, the fairness of the comparison between treatment and control segments – is assessed without knowledge of the consequences that the assessment has in terms of estimating the causal effects of the proposed regulations.

I have had direct experience overseeing such matching processes in other projects. One such project took place at the CDC's Division of Reproductive Health where the task was evaluating the consequences of different guidelines for *in vitro* fertilization. Another was for the Institute for Employment Research at the German Federal Employment Agency, which was evaluating the effectiveness of various job-training programs. Again, the process can be time-consuming, but it ensures that all stakeholders understand and accept that what is being sought is an honest and fair assessment of the proposed interventions.

When buy-in is complete, and the treatment and control segments are considered to be essentially at least as balanced with respect to the relevant covariates as if they had been randomly divided, a protocol for the analysis of the outcome data should then be agreed upon. This step presents researchers with many methodological options, but again, we have substantial experience with such methods. This is expanded upon in Step 5, which follows.

5. **Protocol for the Analysis of Data.** The researchers would assess the efficacy of the FAA's proposal, *i.e.*, whether it would improve flight safety, by comparing and contrasting the matched control and treatment segments in light of the outcome data. Prior to obtaining the outcome data, we would write a protocol describing, as precisely as possible, how outcome data would be analyzed. This process begins by carefully defining quantities to be estimated, known as "estimands." The general task of defining estimands and identifying the appropriate analysis plan to estimate them are critically important steps.

6. **Analysis of the Outcome Data.** Once the protocol is written, outcome data would be appended to the data set of matched control and treatment segments. Then, one would analyze the resulting data according to the protocol, for example, by estimating accident/incident rates for the treated flights and for their matching control flights. If the rates are similar for the treated and matched control flights, then there is no empirical evidence that the proposed regulations will have any effect on flight safety. If the rates are significantly lower in the treated flights (those flights that would have been prohibited under the new rules) than in the matched control flights, then there is evidence that the regulations would have actually eliminated safer flights, *i.e.*, that there are negative effects of the proposed regulations; see the discussion below for some possibilities. If the accident/incident rate is significantly higher in the treated flights than in the matched control flights, then there is empirical evidence that

the proposed regulations would improve flight safety, which would suggest a need to do a follow-on cost-benefit analysis of the regulations.

7. Assessment of Other Possible Consequences of the Proposed Intervention

The data analysis discussed in Step 6 may reveal unintended adverse effects on safety that stem from particular provision(s) of the proposed regulations; this could suggest modifying those provisions if a decision is made to proceed with the regulations. However, that analysis could not fully address all the possible unintended negative consequences of the regulations. This optional Step 7 would use the outcome and covariate data from Steps 1 – 6 to examine such potential consequences.

The risk and gravity of unintended consequences should never be discounted. As an illustrative example, with which I am personally familiar, the National Institute of Health discovered, after extensive study based on randomized experiments, that hormone replacement therapy, long believed to prevent cardiac problems in post-menopausal women, actually worsened them. [6] Another example, from the transportation sector, was recently highlighted by the National Research Council of the National Academies in their new report on airline pilot commuting.[7] Because their discussion is directly related to the question at hand, I quote from the relevant portion of the report in its entirety:

> A major concern in establishing any regulation is designing it so that it achieves its intended effect. Negative unintended consequences often emerge when a seemingly simple regulation is implemented in a complex system. Regulators may not have enough knowledge about the detailed operation of the systems and so may adopt seemingly simple regulations that fail to anticipate how the system will respond to those regulations. An early analysis of the general problem of unintended consequences found

---

[6] An excellent discussion of these findings can be found in *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results from the Women's Health Initiative Randomized Controlled Trial,* Journal of the American Medical Association, Vol. 288, No. 3 (2002), available at www.jama.ama-assn.org.

[7] *The Effects of Commuting on Pilot Fatigue,* National Research Council (2011), available at http://www.nap.edu/catalog.php?record_id=13201

that one of its sources is "imperious immediacy of interest," which is where the intended consequence of an action is desired so strongly that potential unintended effects are purposely ignored (Merton, 1936).

The committee is concerned that a rush to establish regulation regarding pilot commuting and fatigue without an adequate understanding of how pilot commuting and fatigue interact with the aviation system might trigger unanticipated and unintended consequences that have not yet been carefully anticipated.

Such unanticipated and unintended consequences can reduce the effectiveness of the regulation in achieving its goal, and in some cases may even result in a regulation having the opposite effect of what had been intended. A noteworthy example occurred with the 55 mph speed limit, established in March 1974 in response to the 1973 oil embargo (see National Research Council, 1984). Following this adoption, highway fatalities dropped. Although multiple factors contributed to the decline in fatalities, the general consensus was that the reduced speed limits had resulted in fewer highway fatalities. As the fuel shortage eased, the speed limit was retained largely on the grounds of the increased safety it apparently provided. However, in response to other pressures and interests, in 1987 40 states raised the speed limit to 65: many anticipated that highway fatalities would again increase due to the higher speeds. Although highway fatalities did increase, so did vehicle miles traveled. A study that examined statewide fatality rates, considering not only the roads on which the speed limits were changed but also the non-interstate roads on which they were not, found that the higher 65 mph speed limit reduced the statewide fatality rates by 3.4-5.1 percent in comparison with other states (Lave and Elias, 1994). It appears that this unexpected and initially counterintuitive result was because enforcement and highways are integrated systems. The federal government had threatened to impose financial penalties if the 55 mph speed limit was not enforced, so states devoted considerable patrol resources to rural interstates and reduced both enforcement on other highways and other safety activities. In addition, it appears that the higher level of enforcement on rural interstates may have caused some drivers to switch to parallel non-interstate highways, which are more dangerous in terms of fatalities but which had the same speed limit and less speed enforcement. Where the 55 mph speed limit was raised and the threat of federal financial sanctions removed, highway patrols reallocated their activities to a better balance from a safety perspective and with the higher interstate speed limits, some drivers switched from parallel rural roads to the safer interstate highways.

This experience offers two cautionary lessons for safety regulation. One is that complex systems may react to regulation in ways that were unanticipated and, in this case, counter to the goal of improved highway safety. The second is that patterns of enforcement can have important and often unanticipated effects on how a system reacts to regulation.

In aviation safety regulation, another possible unintended consequence of regulation can come from modal shift effects. A regulation that increases costs to the airline industry will likely result in some portion of those costs being passed on to travelers in the form of higher airline ticket prices. Higher airline ticket prices would cause some travelers to switch their mode of travel from airplanes to automobiles. Since travel by private automobile is more dangerous than travel by commercial airline, the result of such a shift would be an increase in highway fatalities. Thus, however many airline passenger and crew lives are saved by the airline safety regulation; the net savings of life from the regulation would be less because of the increase in highway fatalities. In some cases, the net effect may actually be a net loss of lives from a regulation intended to save lives.

This potential for an outcome other than that intended was forecast for a proposed regulation to mandate the use of child safety seats on commercial airlines. One study of the proposed regulation (Windle and Dresner, 1991) concluded that more lives would be lost from the switch to highway travel from the higher travel costs for families with children than would be saved from the added safety benefit of child safety seats.[8]

Although that discussion takes place in the context of assessing the potential side effects of regulating pilot commuting habits, in my judgment their observations are equally applicable to an evaluation of the FAA's proposed rule on crewmember flight, duty and rest requirements.

In our situation, one expected result of the proposed rules (according to public comments) would be the hiring of thousands of additional pilots by the airlines in order to operate legally their current flight schedules. We would need to assess the effects, if

[8] *Id.*, at 6-8 – 6-9.

any, on the incident/accident rate from the introduction of new pilots into the workforce. This would require a different matching process, because the evaluation (outlined in Steps 1-6, above) addresses only the immediate consequence of cancelling the prohibited flights, as opposed to operating those flights with additional crews. Questions for examination would include: As in the National Research Council report, would the net result be an increase in ticket prices such that more people would drive instead of fly? Would the airlines be forced to hire less experienced pilots? In the latter case, we could try to assess the causal effects of pilot experience by matching segments with experienced pilots and inexperienced ones.

Projected Schedule for Completion of the Study

In my experience, I believe that, with the proper resources, the proposed study could be completed within 24 months (two years) of commencement, but this is very dependent on time for interested parties to agree to buy in, the ease of assembling data bases, the ability to create matches, etc; The estimated range of time involved in each phase is as follows:

1. Initial Discussions	2 to 4 months
2. Data Gathering	3 to 5 months
3. Segment Classification	1 to 3 months
4. Segment Matching	5 to 7 months
5. Protocol for the Analysis of Data	1 month
6. Analysis of the Outcome Data	2 to 4 months
7. Assessment of Other Possible Effects	Currently beyond the scope
	Total: 13 to 24 months

Proposed Research Team

Statisticians: D.B. Rubin and others acting under my direction.

It would be necessary to hire a vendor specializing in database creation and management. These firms vary in size from relatively small with dozens of employees to relatively large with thousands of employees. These entities include the Institute for Social Research, Trinity Partners, Westat, National Opinion Research Corporation, MITRE, and RAND. I have worked with each of these vendors in the past.

Estimated Costs

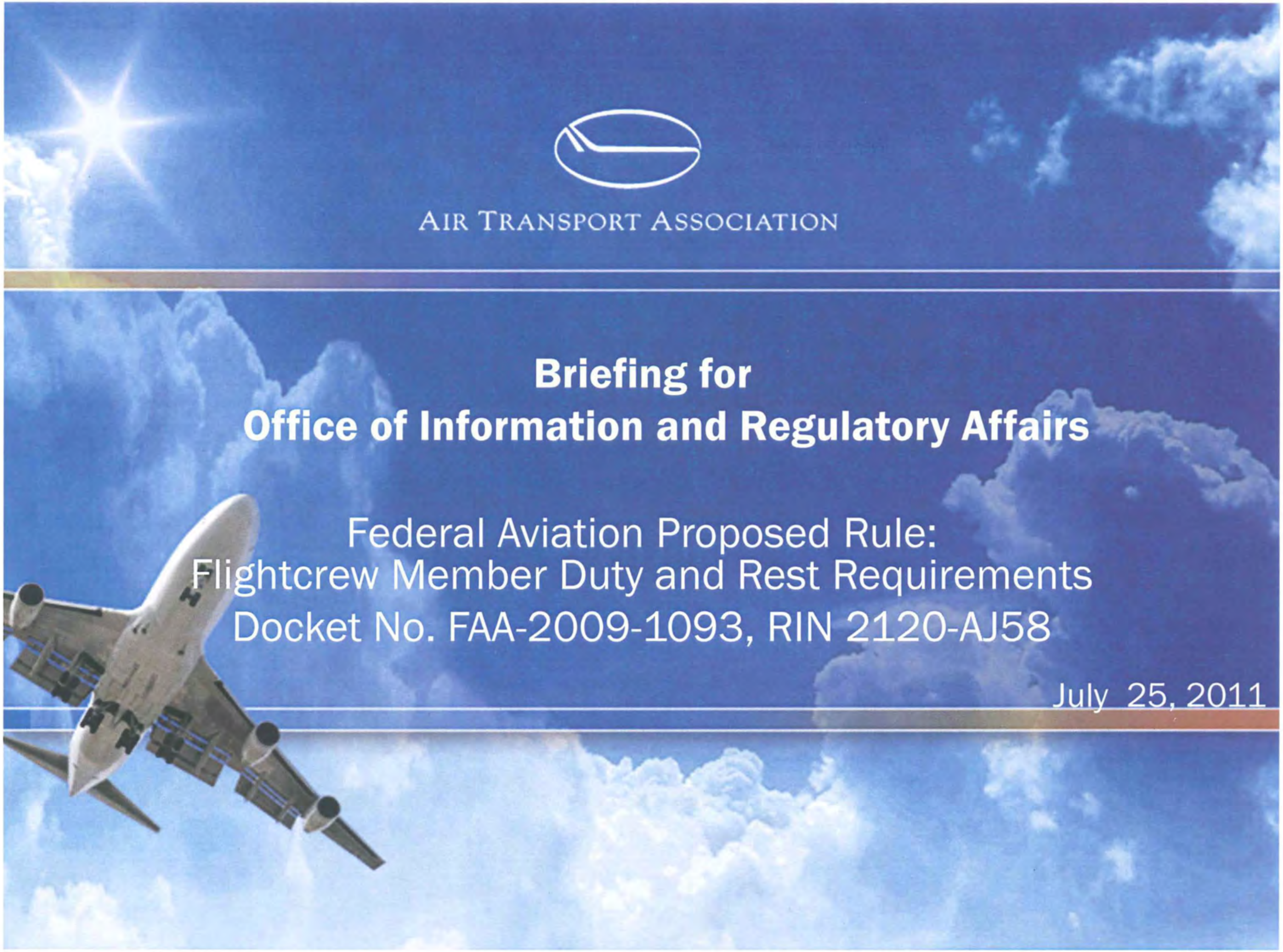
Upon request, we can provide OMB with an estimate of our projected expenses. However, because of the variety of data sets available for the analysis, and the need to identify experts who are familiar with manipulating these data sets and the underlying software systems, it will be difficult to provide a firm estimate until we have completed Step 1, the initial discussions. Those discussions would allow us to determine the relative division of labor between the various stakeholder groups and our research team, as well as outside third parties, to conduct the study.

APPENDIX: SUMMARY OF THE 2010 FAA PROPOSAL

The FAA proposal has been broadly understood as a substantial and historic revision of existing regulations governing flight crew duty and rest. Many flight schedules and crew work schedules permissible today would be rendered illegal. In other cases, schedules that are currently non-viable may become a norm. If promulgated, the regulations would pervasively affect airline operations in term of flight schedules, staffing, route structures, training requirements, and even the types of aircraft airlines choose to operate. The FAA proposal contains several elements that are summarized in the table below:

Types of Airline Operation	The FAA proposal would eliminate historic distinctions between "domestic", "flag", and "supplemental" operations which have had separate rules for flight crew duty and rest. "Domestic" operations are generally those within the 48 contiguous United States. "Flag" operations are generally those occurring between the United States and foreign countries and also between the Hawaii/Alaska and points outside of those states. "Supplemental" operations are generally charter and certain cargo operations. Under the FAA proposal, the same regulations would be applied to all operations for the first time. Thus a pilot flying only regional domestic routes would be subject to the same rules as a pilot operating an intercontinental flight for the first time.
The Concept of "Flight Duty Period" or "FDP"	The FAA historically has regulated how much flight time, or time spent operating the aircraft, a crew member could perform. The FAA proposal would borrow from European regulations and regulate 'duty' periods to encapsulate pre- and post-flight obligations. In contrast to European regulations, the FAA would retain regulations of flight time. Thus the FAA would regulate both flight duty periods and flight time --which has no regulatory precedent. Pilots may run out of flight time with duty time remaining in certain permutations. The legally permissible length of a flight duty period would depend upon the time of day a pilot actually reports to work rather than the originally agreed upon schedule and thus a pilot's FDP could fluctuate overnight. Many flight schedules operated today would become impermissible.
Acclimation Credits and Penalties	The FAA proposal would establish an entirely new system to regulate crew 'acclimation' which would determine which flight duty periods apply. The reference time zone for acclimation would be either the home base or wherever the pilot commences their flying schedule. However it does not take into account that pilots often reside in and report from a time zone other than home base before commencing duty. Many flight schedules operated today would become impermissible.

Flight Time Limitations	Under existing regulations, airlines are permitted to extend pilot flight times to adapt to unforeseen delays caused by weather, air traffic control delay, or other factors. Under the FAA the proposal, airlines' ability to adjust would be restricted. Many flight schedules operated today would no longer be permitted.
Night and Day Operations	The FAA would begin to distinguish between flight operations taking place during daytime and nighttime hours and apply different rules for day and night operations. The FAA would for the first time limit the number of consecutive nighttime operations that may be flown by the same crew. Many crew schedules currently used would become impermissible.
Rest periods	The FAA would increase the minimum required rest by 25% during any seven consecutive day period and would increase minimum daily rest from 8 to 9 hours. For flight segments over certain lengths, the extended rest periods would require changes to flight schedules.
On board rest facilities	The FAA proposes to mandate the installation of different categories of flight crew facilities and regulate flight durations based on the type of facility installed. The effect of the proposal would be to limit or prohibit the types of routes that may be legally operated under existing regulations.
Schedule reliability	The FAA would require airlines to adjust crew duty periods within 60 days if actual duty periods exceeded scheduled duty periods more than 5% of the time any scheduled flight duty period exceeds the schedule 20% of the time. This would be an entirely new requirement which would render certain existing flight schedules impermissible.
Reserve pilots	The FAA would require the use of a new system of scheduling pilots for reserve duty during which they are 'on call' to substitute for crew unable to fly due to illness, missed connection, or other reason. The proposal would create different categories of reserve duty such as "airport/standby, short-call, and long-call" which would be new.


AIR TRANSPORT ASSOCIATION
Briefing for
Office of Information and Regulatory Affairs
Federal Aviation Proposed Rule:
Flightcrew Member Duty and Rest Requirements
Docket No. FAA-2009-1093, RIN 2120-AJ58
July 25, 2011

Overview

- FAA's stated purpose is to rewrite longstanding operational flightcrew member duty and rest requirements to mitigate fatigue
- ATA members share the FAA's goal but any change must be based on operational and safety data and science. Our safety record, active participation in the Commercial Aviation Safety Team, and the Flight and Duty Time Aviation Rulemaking Committee reflects our commitment to safety.
- We cannot support the proposal as written.
  - It includes items with high cost and operational impacts not related to safety and
  - Fails to incorporate science-based concepts as described by fatigue and sleep experts.
- The FAA's safety analysis and benefit-cost analysis were flawed. From a safety and benefit perspective, the agency did not provide specific data directly linking new provisions to areas of concern. The highest impact cost provisions were unrealistically minimized and whole categories of cost items were excluded.
  - Actual benefit likely to be 40% less than FAA estimate
  - Actual cost nearly 16 times FAA estimate
- Even without correcting the BCA, <u>the FAA determined that costs would outweigh the benefits</u>.
- The proposal needs to be vastly improved and revised under Executive Order 13563 and 12866 principles and requires a high level of scrutiny and review before proceeding.
- OMB should ensure FAA provides a "reasoned determination" that benefits outweigh costs seeking to improve the actual results of regulatory requirements. These goals can only be met with substantial revision to the proposal and a renewed effort to accurately measure the impact of any final rule and issuance of a Supplemental Notice of Proposed Rulemaking.

Why a Supplemental NPRM is Needed

- Any FDT proposal must consider all segments of the aviation industry and decades of safe operational experience under current FAA regulations.
- Schedule reliability, flight time limits, and prescriptive extensions to scheduled FDPs impose high costs without mitigating fatigue and should be removed
- FAA should focus on three core elements for mitigating fatigue: daily FDP limits, cumulative FDP limits, and minimum rest requirements
    - EASA issued a drastically different FDT proposal that focused on these core elements
- A well-developed FRMS scheme should also be included to recognize existing fatigue mitigation measures – any FRMS scheme must be fully mature, clear, well defined and ready to implement
- Even if the agency corrects shortcomings or removes unjustified provisions:
    - The public should have an opportunity to review and comment on new justifications/safety benefits in a supplemental proposal
    - Any changes to correct the original proposal to meet EO 12866 and EO 13563 requirements would have to be so dramatic as to prevent:
        - "an open exchange " of information among government officials, experts, stakeholders, and the public (M-11-10, p2; EO 13563, Section 2(a))
        - Input from "the views of those who are likely to be affected" (M-11-10, p2; EO 13563, Section 2(c))
        - "the opportunity to react to (and benefit from) the comments, arguments, and information of others during the rulemaking process" (M-11-10, p2)

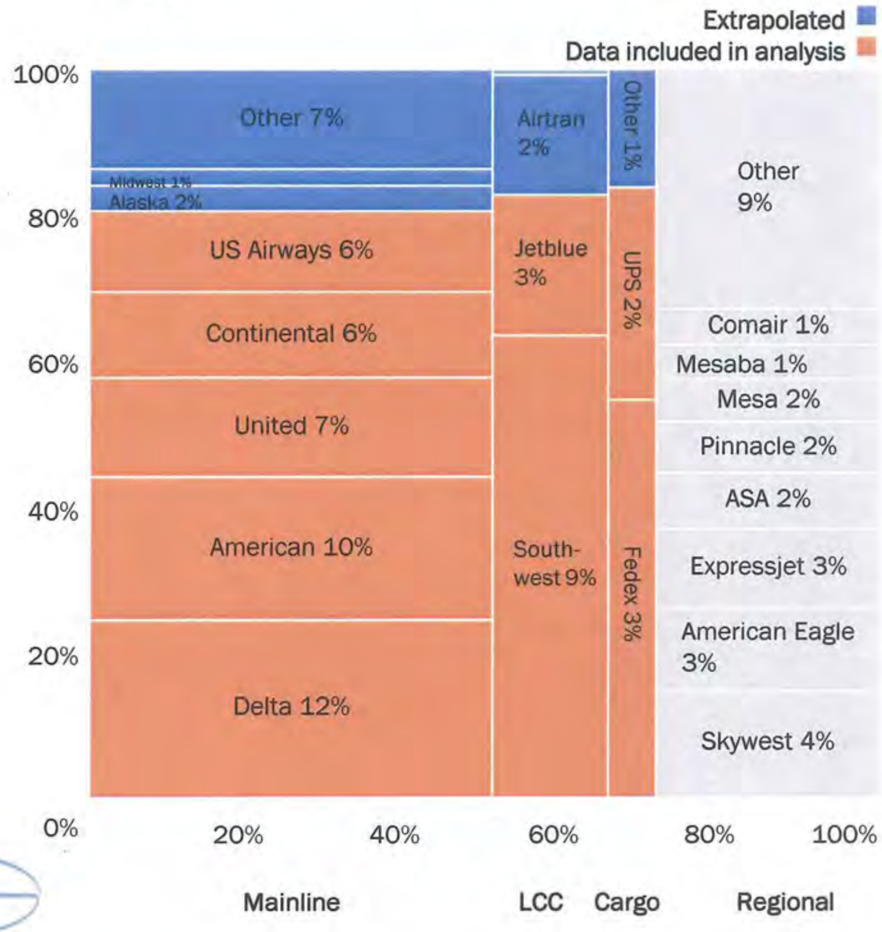
Analysis Conducted

Types of Analysis Conducted

- Reviewed FAA conclusions regarding accidents it cited as the source of its benefits analysis
- Reviewed other FAA assumptions, e.g., flight crew costs, optimization, reduced absenteeism
- Schedule and cost modeling using carrier crew schedules and rostering data. Representative airlines provided raw data for Oliver Wyman to assess the impact on mainline hub and spoke, low cost carrier, and cargo carrier segments
- Aggregation of individual carrier estimates and analyses for some categories of analysis, such as the cost of training and crew rest infrastructure

Unlike FAA "black box" approach, Oliver Wyman assumptions are clearly stated and its modeling results can be replicated by a 3rd party. We welcome a validation of these findings.

Industry Distribution of block hours 2009



FAA Economic Analysis Overstates Benefits and Grossly Understates Costs

- Actual benefit likely to be 40% less than FAA estimate
- Actual cost nearly 16 times FAA estimate
- Actual cost includes only impact to mainline LCC, large cargo and nine of fourteen cost items

10-year Nominal Cost ($ millions)		
	FAA Regulatory Impact Analysis	Oliver Wyman Analysis*
Benefits	$659.4	$395.6
Costs	$1,254	$19,641
Cost/Benefit Ratio	1.9:1	50:1

10-year NPV Cost ($ millions)		
	FAA Regulatory Impact Analysis	Oliver Wyman Analysis*
Benefits	$463.8	$278.3
Costs	$803.5	$14,439
Cost/Benefit Ratio	1.7:1	52:1

* Oliver Wyman NPV calculation uses same discount rate as FAA, 7%

Why did the FAA misestimate benefits and costs?

Benefits –

- FAA labeled accidents as fatigue-related even when NTSB found otherwise
- FAA did not apply its own stated methodology of screening accidents, instead pulling in unrelated accidents from outside its database

Costs –

- By omitting the cost of schedule buffering required by multiple provisions of the NPRM, the FAA omitted the major source of cost to the industry
- FAA "assumed away" other important cost impacts, particularly in areas where it lacked modeling capabilities
- FAA used unrealistically low labor costs
- FAA makes clear that its cost estimates only include those related to individual flight duty periods and does not include the substantial costs incurred as a result of the impact of new duty limits imposed over longer periods of a week or a month

Benefits Estimated by FAA Are Substantially Overstated – 40% of the Accidents Cited Were Improperly Classified

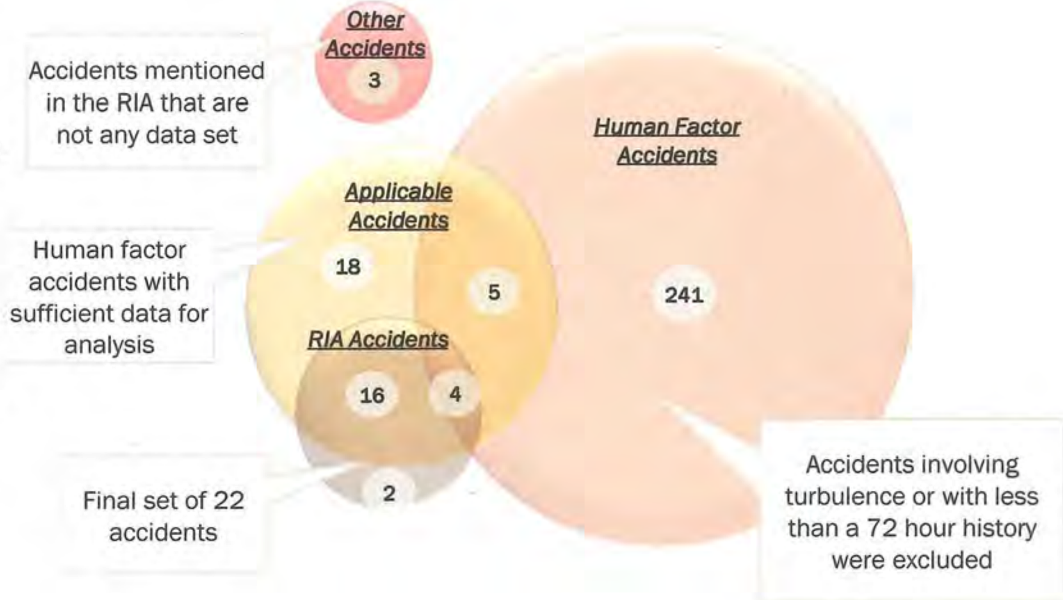
FAA methodology

- **FAA begins with set of 250 accidents, which it then winnows down to 43 accidents that are "human factors" related and a smaller set of 22 accidents, which it concludes are "fatigue-related" with sufficient data for analysis**
- **From these 22 accidents, the FAA extrapolates and projects future accident avoidance benefit results**

Problems with FAA analysis:

- **Contrary to FAA report, the 22 accidents are not a subset of 43 accidents, which are not a subset of the 250 accidents. At a minimum, the chain of analysis is broken and cannot be replicated.**
- **Actual FAA dataset from which it has drawn conclusions has 20 accidents, not 22**
- **Of these 20 accidents, which the FAA cites as avoidable under the NPRM, 40% should be excluded because they have nothing to do with pilot fatigue or the type of flying regulated by the NPRM**
- ***In 3 of the 20 cases, NTSB specifically determined that fatigue was not a factor***
- **Numerous other analytic problems, as explained in the report**
- **Campbell Hill report delves into the accident analysis in greater detail**

FAA Accident Analysis Sets – The FAA reported that each smaller set is a subset of the original set of 250 accidents, but that is not the case



Costs Estimated by FAA Are Grossly Understated – Actual Costs Are at Least 15 Times Higher

Most Costly Items

- **Schedule reliability – Proposed rule would require airlines to add substantial buffers to their flights to meet new 95% schedule reliability standard. [Not found in EU or elsewhere; would not reduce fatigue]**
- **Flight duty period extensions – Proposed rule severely limits FDT schedule extensions – even when those extensions still would be within the new flight duty limits. [Not found in EU or elsewhere]**
- **Flight time limits – Proposed rule incorporates both block (actual flying time) limits and flight duty limits, and makes the block hour limit much more restrictive than current rules by prohibiting extensions to accommodate day-of-operation delays. [Not found in EU or elsewhere]**

Cost estimates for these three provisions:

- **Oliver Wyman estimate: 10-year cost $15.7 billion for mainline, LCC, and large cargo industry sectors**
- **FAA estimate: $765 million, which also includes the cost of other Flight Operations provisions.**
- **Note: FAA estimate of schedule reliability provision only includes the cost of monitoring schedule reliability**

	FAA RIA	Oliver Wyman Analysis
Fatigue Training	262	331
Crew Rest	227	928
Schedule Reliability	5	9,624
Flight Operations	760	8,758
Total	$1,254	$19,641

* Flight Operations bar in graph includes flight time limits, flight duty period extensions, and other provisions with much lower costs such as minimum rest between duties, day of operations reserve,

Full Cost of NPRM Substantially Exceeds Even Oliver Wyman Estimate

Extrapolating to Include Regionals

- **Oliver Wyman cost estimate includes only mainline, LCC, and large cargo**
- **FAA NPRM includes, in addition, regional, small cargo, small passenger, and charter passenger costs**
- **Extrapolating the cost of relevant provisions to regional carriers (the largest segment not included in the OW analysis) add $1.987 billion in costs**

The Full Cost of Schedule Reliability Provisions

- **As explained in the report, the OW cost estimate for the schedule reliability rule is based on the most flexible/favorable interpretation of the proposed rule. The $9.6 billion cost estimate may understate by as much as $50 billion the true cost of complying with the rule as written**

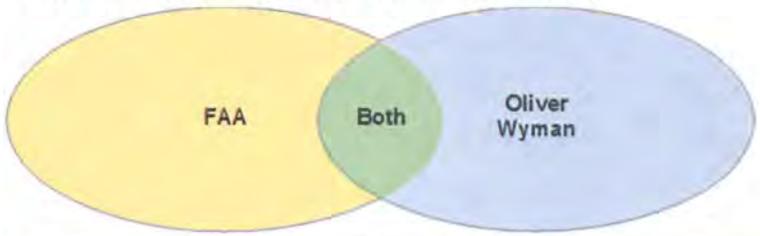
Including Other Cost Categories

- **As explained in the report, given the short time period to respond to the NPRM, the Oliver Wyman analysis includes some but not all cost impacts. E.g., it excludes the five cost categories listed in the orange box to the right**
- **A more complete analysis would also include the cost of rest requirements and several other provisions**

NPRM Cost Estimate

10-Year Nominal Cost (Including Regional Carriers)

Provisions Analyzed by Oliver Wyman and FAA (RIA)



FAA *Topic area not included in Oliver Wyman report*	Both *Topic area appears in both reports*	Oliver Wyman *Topic area not included in FAA report*
• Rest Requirements • Flight Duty Period – Minimum Rest Between Duties – Reduced Augmentation – FMRS Development Costs – Flight Engineer Supplemental Operations	• Flight Time Limits • Schedule Reliability • Fatigue Training • Crew Rest Facilities	• Flight Duty Period – Day of Operations Reserve – Cumulative Duty Time from Short Call Reserve – FDP Extension – Three Consecutive Nights – Collective Bargaining Agreements (partially included in RIA analysis)

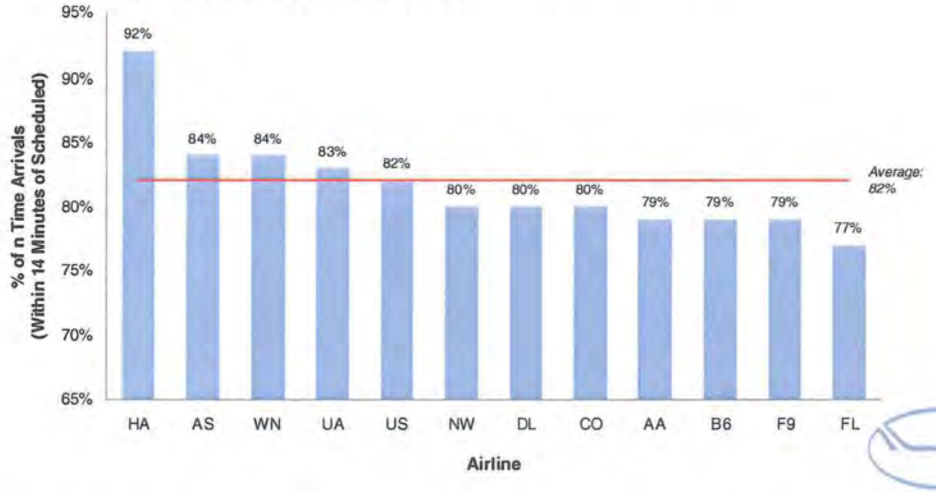
Other Deficiencies with FAA Regulatory Impact Analysis

Subject	Comment
5% reduction in absenteeism assumed to result from NPRM	No basis for FAA assumption
Collective bargaining agreement impacts on the costs/benefits of the rules	FAA assumes CBAs will be adapted to match rules without any cost – not realistic
25% optimization assumption	FAA assumes that airlines will find a way to optimize implementation so as to save 25% of the estimated implementation costs – no basis
Cumulative impacts	FAA specifically excludes whole categories of analysis: "Only limits relating to individual flight duty periods were applied. Cumulative limits were not applied due to data limitations."
Flight crew costs	FAA appears to have used raw average salary data, without payroll taxes, pension, and benefits – which substantially understates true costs
Cancellation, buffer, and delay costs	FAA assumes that all provisions, regardless of how restrictive, can be implemented without the carriers incurring any cancellations, delays, or adding schedule buffers but provides no other means to meet "hard time limits
Actual versus scheduled performance	FAA prohibits carriers from extending scheduled flight duty times and scheduled flight times even under circumstances when the extended times would be well under the proposed maximums. This requirement (unique in international safety regulations) appears to add enormous costs without aiding safety or reducing fatigue

Examples of Specific Provisions – Schedule Reliability

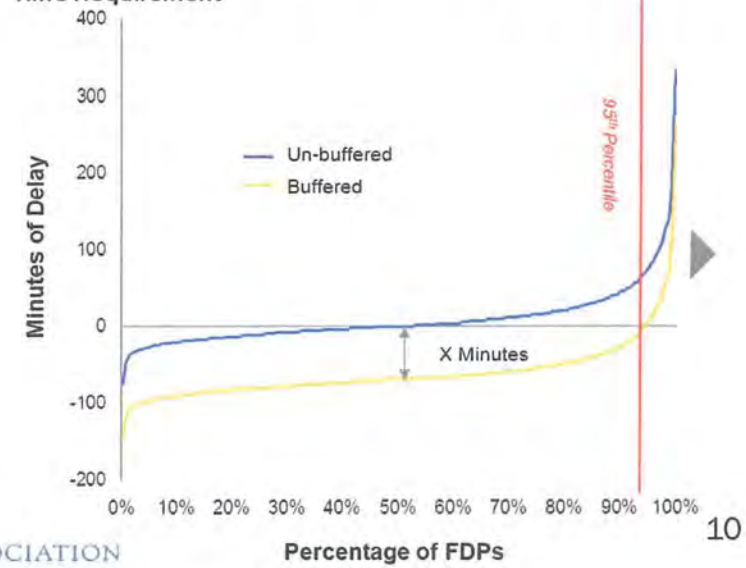
- NPRM requires carriers to achieve 95 percent scheduled/actual Flight Duty Period rate, which means 95% of flights must arrive as scheduled with 0 minutes tolerance for lateness
- Currently, most airlines are structured to achieve between 50% and 60% actual vs planned
- To achieve 95 percent FDP rate (+/- 0 minutes), airlines will need to add schedule buffers (i.e., add block time)
  - Due to the lack of delay predictability all FDP's will need to be buffered
- This will substantially impact airline costs without improving safety or reducing fatigue:

Current 82% average on-time arrival rate (0-14 minutes) is equivalent to 50-60 % on-time (+/-0)



Source: US DOT Air Travel Consumer Reports 2009

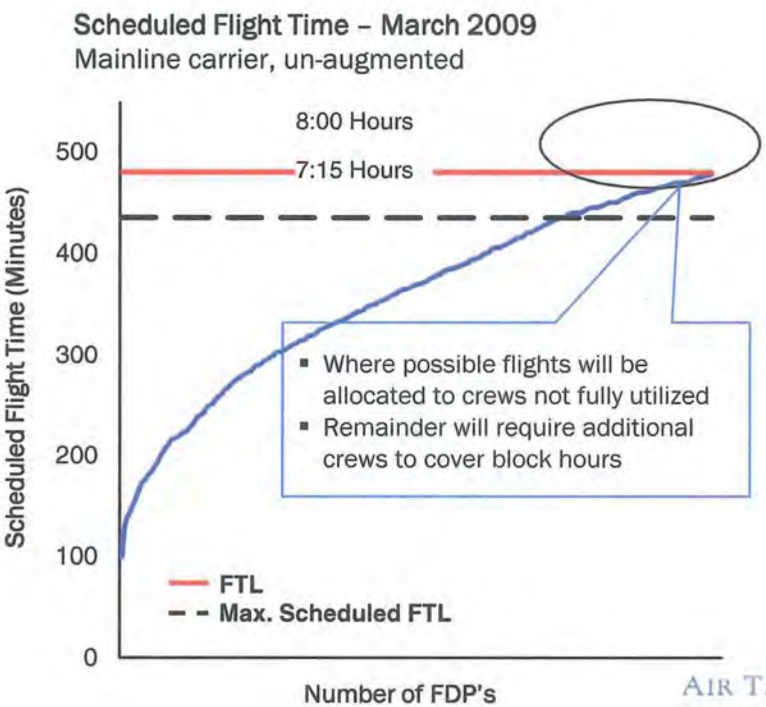
Schedule Adjustment Required to Meet 95% FDP Schedule On-Time Requirement



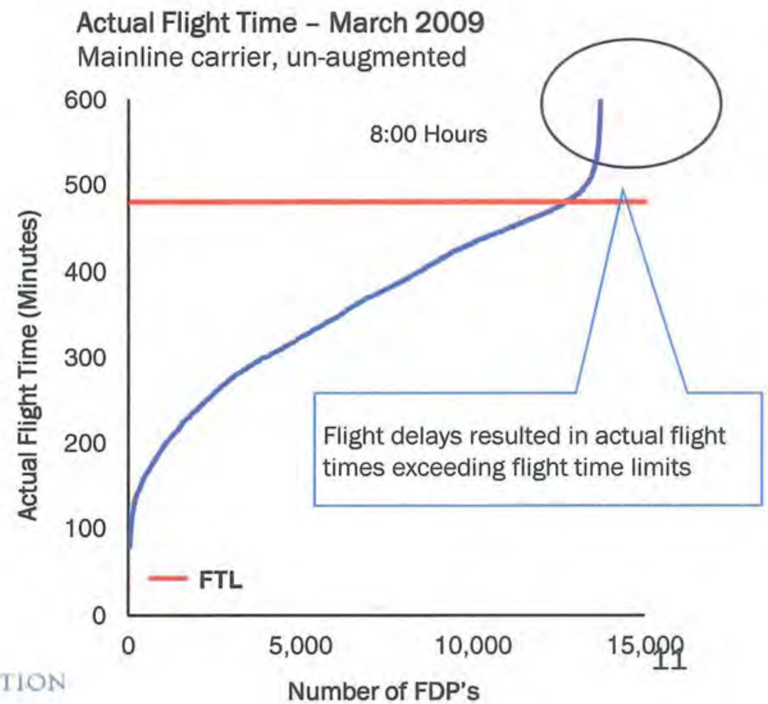
Examples of Specific Provisions – Flight Time Limits

- Current rule permits flight crew to operate a flight when due to circumstances beyond the carrier's control (such as adverse weather), the flight is not expected to reach its destination within the scheduled time.
- NPRM, however, makes flight time limits inflexible "hard limits" *even when total flight duty time remains under the new maximum*
- As a result, carriers must schedule for well under the flight time limits, and also must incur the cost of cancellations that occur as a result of delays beyond their control pushing flight time beyond the flight time limit

Must schedule well under flight time limit



Must cancel when bad weather results in hitting limit



AIR TRANSPORT ASSOCIATION

Cost Summary

NPRM Provisions	Impact (Millions, Assuming 25% Optimization)	
	Oliver Wyman	FAA Regulatory Impact Analysis
Flight Time Limits*	10 Year Cost: $15,740 (Individual estimates provided in Report; these three provisions are interrelated, and cost of each depends on allocation assumptions)	10 year cost: $760 (Part of flight operations)
Schedule Reliability*		10 Year Cost: $5
FDP Extension*		Not quantified
Day of Operation Reserve	10 Year Cost: $826	Not quantified
Cumulative Duty Time from Short Call Reserve	10 Year Cost: $143	Not quantified
Crew Rest Infrastructure	10 Year Cost: $928	10 Year Cost: $227
NPRM Implementation	10 Year Cost: $1,967	10 Year Cost: $262
Three Consecutive Nights	10 Year Cost: $38	Not quantified
Totals (10 Year Additive)	$19,641 Nominal	$1,254 Nominal

Cost Summary after Removal of Schedule Reliability and Flight Duty Period Extension Provisions

NPRM Provisions	Impact (Millions, Assuming 25% Optimization)	
	Oliver Wyman	FAA Regulatory Impact Analysis
Flight Time Limits	10 Year Cost: $4,280	10 year cost: $760
Day of Operation Reserve	10 Year Cost: $826	Not quantified
Cumulative Duty Time from Short Call Reserve	10 Year Cost: $143	Not quantified
Crew Rest Infrastructure	10 Year Cost: $928	10 Year Cost: $227
NPRM Implementation	10 Year Cost: $1,967	10 Year Cost: $262
Three Consecutive Nights	10 Year Cost: $38	Not quantified
Totals (10 Year Additive)	$8,182 Nominal	$1,254 Nominal

EASA FDT Proposal

The European Aviation Safety Agency issued a flightcrew member Flight and Duty Time proposal this year that was drastically different from FAA:

- EASA did not include daily flight time limits
- Schedule Reliability is presented as guidance in EASA regulations
- FDP extensions are to maximum limits not to a flightcrew member's schedule
  - Preplanned extensions are limited to one hour over maximums, day of extensions are limited to two hours over maximums
- EASA flightcrew member rest periods remain the same
  - EASA: Minimum of 12 hours of rest at a pilot's home base, and 10 hours rest when away from a home base
  - FAA: currently has an 8 hour rest requirement and proposed a 9 hour rest requirement
- EASA proposal is much more flexible because it concluded:
- "the assessment of safety impacts for this RIA could not be based on statistical data from accidents and incidents as there was no statistically significant number of accidents and incidents for EASA-country operators."
- EASA focused on core areas directly related to fatigue as the FAA should

Executive Orders

Executive Orders 13563 and 12866 set out a number of regulatory principles, to which the FAA has not adhered in this rulemaking.

- **First, "[f]ederal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling public need...."**
  - **The FAA safety analysis did not demonstrate a compelling public need for the proposal. FAA should not include highly burdensome regulations with many provisions not related to safety and include only such provisions that directly address areas of concern.**
- **Second, "[e]ach agency shall assess both the costs and the benefits of the intended regulation and ... propose or adopt a regulation only upon a reasoned determination that the benefits of the intended regulation justify its costs."**
  - **The FAA's assessment of the costs and benefits of the NPRM was deeply flawed, and did not represent a "reasoned determination" because *the agency concluded that costs would outweigh benefits.***
- **Finally, "[e]ach agency shall base its decision on the best reasonably obtainable scientific, technical, economic, and other information concerning the need for, and consequences of, the intended regulation."**
  - **FAA admits that aspects of the proposal are not supported by science**
  - **FAA ignored highly relevant technical, economic, and operational information**

ATA members and outside fatigue and economics experts agree the FAA's proposal was flawed. For these reasons, the NPRM did not meet E.O. 13563 and 12886 standards, and should be <u>substantially reviewed and revised before proceeding to a Supplemental Notice of Proposed Rulemaking.</u>

Essential Changes Needed for a SNPRM

1. **The Final Rule should continue to recognize different operational models**
   - Any final rule should recognize and respond to different air carrier operational environments and models, including domestic and international passenger operators, domestic and international cargo operators, and on-demand (nonscheduled) charter operators.
   - Nothing in fatigue/sleep research suggests a need for a one-size-fits-all regulation.
   - Science-based guidelines, judiciously blended with many years of operational experience, will allow the various air carrier models to continue to operate safely.
2. **Remove proposed schedule reliability requirements, they have nothing to do with safety**
3. **Remove daily flight time limits, no other regulatory scheme in the world uses these limits**
   - Rest requirements, cumulative flight time limits, and daily and cumulative flight duty periods mitigate fatigue
4. **Allow FDP extensions to actual operations, eliminate the NPRM proposal limiting extensions to scheduled FDPs**
5. **Increase minimum rest requirements to 10 hours**
6. **Include a more fully developed FRMS program with clear standards, based on ICAO principles, years in advance of a FDT final rule effective date that carriers can rely on to satisfy new requirements**
7. **Permit "split duty" rest on the ground for a minimum of 2 hours**


AIR TRANSPORT ASSOCIATION
17