



Equal to or better than: The application of statistical non-inferiority to fatigue risk management

Amanda Lamp*, Jane M.C. Chen, David McCullough, Gregory Belenky

Sleep and Performance Research Center and Elson S. Floyd College of Medicine, Washington State University, 412 E. Spokane Falls Blvd., Spokane, WA, 99202, United States



ARTICLE INFO

Keywords:

Aviation
Non-inferiority
Fatigue risk management
Sleep
Pilots
Human performance

ABSTRACT

In December 2014, the Federal Aviation Administration (FAA) completed a major revision of the rules and regulations governing flight and duty time in commercial aviation (Federal Aviation Regulation (FAR) Part 117). Scientists were included in the revision process and provided insights into sleep, sleep loss, the circadian rhythm, and their effects on performance that were incorporated into the new rule. If a planned flight was non-compliant with the regulation, for example if it exceeded flight and duty time limits, it could only be flown under an FAA-approved Fatigue Risk Management System (FRMS) as meeting an Alternative Method of Compliance (AMOC). One method that a flight could qualify as an AMOC is if it could be demonstrated empirically that it was as safe as or safer than a similar flight, designated the Safety Standard Operation (SSO), that was compliant with the regulation. In the present paper, we demonstrate the FRMS process using a comparison between a non-compliant AMOC flight from the US west coast to Australia and a compliant SSO flight from the US west coast to Taiwan. The AMOC was non-compliant because it exceeded the flight time limits in the prescriptive rule. Once a data collection exemption was granted by the FAA, both the outbound and inbound AMOC and SSO routes were studied on four Safety Performance Indicators (SPIs). The SPIs studied were inflight sleep, cognitive performance, self-reported fatigue, and self-reported sleepiness. These measures were made at top of descent (TOD), a critical phase of flight. The study was designed as a paired comparison. Forty volunteer pilots studied flew both the AMOC and the SSO flights for a total of 80 studied flights. Using statistical non-inferiority applied to the AMOC and SSO SPIs, we demonstrated, as required by the new rule, that the US-Australia AMOC flight was “as safe as, or safer than” the US-Taiwan SSO flight. In the context of FRMS, statistical non-inferiority is a concept and technique of great utility, straightforward in application, producing clear visual representations of the findings, and providing a direct answer to the question posed by the regulation - is the AMOC flight “as safe as, or safer than” the SSO.

1. Introduction

The United States Federal Aviation Administration (FAA), 2013 Federal Aviation Regulation (FAR) 14 CFR Part 117 is a major revision of the regulations governing flight and duty time in commercial aviation. This revision integrates scientific findings with respect to sleep, sleep loss, fatigue, circadian rhythm, and performance into the prescriptive rules governing flight and duty-time limits (Flight and Duty Limitations and Rest Requirements: Flightcrew Members, 2012). In an innovative move, FAR subsection Part 117 § 117.7 provides for a data collection-enabled Fatigue Risk Management System (FRMS) as an alternative to the “one-size fits all” prescriptive approach. Section 117.7 stipulates that no operator “may exceed any provision of this part unless approved by the FAA under a Fatigue Risk Management System

that provides at least an equivalent level of safety against fatigue-related accidents or incidents as the other provisions of this part” (Flight and Duty Limitations and Rest Requirements: Flightcrew Members, 2012). When a proposed flight does not fit within the prescriptive envelope, an FRMS can serve as an Alternative Method of Compliance (AMOC) if it can be shown by data collected under an exemption to be as safe as or safer than a similar route that is in compliance with the regulation. At its core, an FRMS is a data-acquisition process to monitor and manage fatigue risks in accordance with the International Civil Aviation Organization’s (ICAO) Annex 6 guidance (2008). In practice, an FRMS is developed jointly with input from sleep and performance scientists, airline management, pilots’ union representatives, and the FAA.

When applied to commercial aviation, an FRMS may include the

* Corresponding author.

E-mail addresses: alamp@wsu.edu (A. Lamp), jane.chen@wsu.edu (J.M.C. Chen), d.mccullough@wsu.edu (D. McCullough), belenky@wsu.edu (G. Belenky).

development and incorporation of mathematical models optimizing performance into rostering and scheduling software (Rangan et al., 2013). Within the broader context of rostering and scheduling optimization, route-specific studies may be undertaken to compare a new route that does not fit within the prescriptive rule, designated the AMOC, to a similar route that is compliant with the prescriptive rule, the Safety Standard Operation (SSO). We will be discussing the application of an AMOC compared to an SSO. The FAA requires the operator to demonstrate that the AMOC is “as safe as, or safer than” the SSO (AC 120-103A – *Fatigue Risk Management Systems for Aviation Safety*, 2013) by collecting and analyzing inflight data on sleep, cognitive performance, and self-reported fatigue and sleepiness. Statistical non-inferiority analysis is used to demonstrate that the designated AMOC route is non-inferior, i.e., as safe as, or safer than the designated SSO route, as required by FAA regulation.

An example of the FRMS process is described below. In the example, the non-compliant AMOC route is from the US west coast to Sydney, Australia (US-Australia) and return. The compliant SSO route is from the US west coast to Taipei, Taiwan (US-Taiwan) and return. All flights have four pilots (double augmented crew). All four pilots are present on the flight deck during take off to top of climb (TOC) and from top of descent (TOD) to landing. During cruise (from TOC to TOD), only two of the pilots are required on the flight deck. When not required on the flight deck, FAA approved Class I rest facilities (bunks) are available for pilot rest or sleep. Given a 16-hour flight, subtracting an hour from each end, and dividing up the available time, each pilot will have approximately 7-hours of sleep opportunity during cruise.

1.1. Initial safety case

Before flying a new route under an FRMS data collection exemption, an FRMS application must be submitted to the FAA that includes an initial safety case for the route. This initial safety case can be based on prior data from similar routes, on mathematical modeling that predicts minute-by-minute performance of the pilots flying that route, or both (AC 120-103A - *Fatigue Risk Management Systems for Aviation Safety*, 2013, Appendix 2 p. 26–27). The two process mathematical model predicts performance based on time awake (sleep loss) and time of day (circadian rhythm) (Achermann and Borbély, 2003; Borbély, 1982; Borbély et al., 2016). The model is used to predict the effects on performance of specific sleep/wake and duty periods a pilot experiences, to estimate the level of degradation in performance due to fatigue, and to provide an estimate of schedule-induced fatigue risk. In order to proceed, modeling performed in support of an FRMS application must indicate an equal or superior level of safety of the AMOC route compared to the SSO route. Modeling can indicate if the AMOC flight is likely to be determined safe (as safe as or safer than the SSO), at critical phases of flight such as TOD. If this initial safety case suggests that the AMOC route is likely to be safe to fly, then the FAA will issue an exemption authorizing the airline to collect data from pilots during actual FRMS operations of the AMOC and SSO flights (AC 120-103A, 2013, Appendix 2 p. 43). This begins the data collection phase.

1.2. Data collection

Four Safety Performance Indicators (SPIs) including inflight sleep, cognitive performance, self-reported fatigue, and self-reported sleepiness are studied on both the AMOC and SSO routes. These SPIs are typically measured at TOD. Even if the FRMS exemption is requested only for either the outbound or inbound segment of the flight, data are collected on both the outbound and inbound flights to analyze risk across the entire flight duty period.

Throughout the data collection period, the data reporting content and frequency is defined by the exemption. This usually includes comprehensive monthly reports including the AMOC and SSO data for TOD which are sent to the FAA. Each monthly report includes the

average of each study measure from all months of collected data, and the average of each study measure for the previous month only. These reports are used to ensure that the AMOC trip does not show unsafe levels of sleep, fatigue, sleepiness, and/or performance during the data collection period.

1.3. Data analysis by non-inferiority

With the standard t-test, failure to show a difference does not equate to equivalence (Greene et al., 2008). A non-significant result may simply be due to a lack of statistical power, and therefore is not a positive indication of equivalence (Mascha and Sessler, 2011). In contrast, a non-inferiority design can be used to test for equivalence, superiority, and non-inferiority. Non-inferiority is a graphical, mean and confidence interval-based approach to statistical superiority, equivalence, and non-inferiority (Walker and Nowacki, 2011). This approach is used in the pharmaceutical industry as well as other medical device development and procedure research, where researchers seek to demonstrate that a new drug, intervention, or technique is as safe as or safer than (equivalent and/or superior to) the one in current use (Greene et al., 2008). In our example, we examine by non-inferiority analyses whether the AMOC route is as safe as or safer than the SSO route.

1.4. Approval for exemption

Once the data has been collected, the data is analyzed with non-inferiority testing and a final report is provided to the airline and the FAA. The FAA reviews all of the information from the application, monthly reports, any fatigue reports sent to the FAA by the airline, and the final non-inferiority analysis report. If the FAA decides that the data and information provided demonstrates the AMOC is safe to fly (as safe as or safer than the SSO), authorization to operate under the parameters of the FRMS is issued via Operations Specifications to the airline for the AMOC route.

1.5. Continuing review

After the FRMS authorization has been issued, continuing reviews occur quarterly to ensure the trip continues to demonstrate the same level of safety (AC 120-103A, 2013, Appendix 2 p. 56). This continuing review includes safety reports and some form of data collection such as including only self-reported fatigue. As the route continues to be flown safely over the years, the continuing review will become bi-annual and then annual (AC 120-103A, 2013, Appendix 2 p. 56).

2. Material and methods

To develop the initial safety case in the present study, the AMOC US-Australia route and SSO US-Taiwan route were modeled and compared using the two-process Sleep, Activity, Fatigue, and Task Effectiveness – Fatigue Avoidance Scheduling Tool (SAFTE-FAST) model (Hursh et al., 2004).

In our example, data collection was performed as a paired comparison. Forty pilots who volunteered for the study flew both the AMOC flight and the SSO flight, for a total of 80 studied flights. Data collection included measures on the four standard FRMS SPIs: inflight sleep, cognitive performance, self-reported fatigue, and self-reported sleepiness. Potential volunteers were recruited through a recruitment message sent out by airline and union representatives. The message notified pilots of their eligibility for the study, briefly explained the study, outlined compensation, and explained that participation and withdrawal were voluntary. If a pilot was interested in volunteering, then they were asked to contact Washington State University (WSU) directly. All volunteers gave informed consent. Study and informed consent documents were approved by the WSU Institutional Review Board (IRB).

The study period began three days prior to each studied flight and ended three days after each studied flight returned. This study period of eight- to nine-days for each of the AMOC and SSO flights allows for a full picture of sleep, performance, fatigue, and sleepiness across time. Sleep was measured continuously throughout the data collection period, while the other SPIs were collected at outbound and inbound TOD. Sleep was measured objectively using a wrist worn actigraph (Phillips Respironics Actiwatch Spectrum) and by self-report using a logbook. An actigraph uses wrist movement to algorithmically estimate when the wearer is awake or asleep (Ancoli-Israel et al., 2015).

The psychomotor vigilance task (PVT) was used to measure cognitive performance. For our study, we used PVT mean speed which is computed as mean (1/reaction time). The PVT is a simple reaction-time task that is sensitive to attentional lapses, with higher speed scores indicating faster, better performance. The PVT has been shown to be correlated with sleep loss (Belenky et al., 2011). The ten-minute version is sensitive to sleep loss and circadian phase with no evidence of learning effects (Gander et al., 2013b; Lamond et al., 2005; Thorne et al., 2005). The shorter five-minute version correlates with the ten-minute version, is recommended by International Civil Aviation Organization (ICAO) (2008), and is currently used in all FRMS aviation studies (Gander et al., 2013b). Research indicates that the PVT is insensitive to learning effects, making it a reliable measure suitable to be used repeatedly throughout the study period. The present study uses the five-minute version of the PVT.

Self-reported fatigue was assessed using the Samn-Perelli Fatigue Scale (SP) and self-reported sleepiness was assessed using the Karolinska Sleepiness Scale (KSS). The SP is a seven-point scale ranging from “1-fully alert, wide awake” to “7-completely exhausted, unable to function.” The KSS is a nine-point scale ranging from “1-extremely alert to 9-extremely sleepy, fighting sleep.” Of note, SP scores of 6–7 have been associated with severe fatigue (Gander et al., 2013a; Samn and Perelli, 1982) and KSS scores of 7–9 have been associated with microsleeps (Åkerstedt and Gillberg, 1990; Gander et al., 2013a). For the purposes of the FRMS, cognitive performance, self-reported fatigue, and self-reported sleepiness measures taken at TOD were analyzed as well as in-flight sleep.

2.1. Theory/calculation

In the language of formal logic, non-inferiority is the logical union of equivalence and superiority. The null and alternative hypotheses for non-inferiority, equivalence, and superiority have been provided in Table 1 for greater explanation into how the three study designs relate to one another. In our example, demonstrating that the US-Australia AMOC is non-inferior to the US-Taiwan SSO on all four SPIs is the same as demonstrating that the US-Australia AMOC is as safe as or safer than the US-Taiwan SSO. Results are displayed on non-inferiority plots and p-values may also be included (Walker and Nowacki, 2011).

Graphical non-inferiority is a statistical approach in which the presence or absence of equivalence, superiority, and non-inferiority can be read directly from the plot without further computation. Initially, in aviation FRMS, equivalence testing in conjunction with difference testing was used to approximate “equal to or better than.” A problem with this approach is that a result may be found where there is neither

Table 1
Comparisons of the null and alternative hypotheses for non-inferiority, equivalence, and superiority where H_0 = Null Hypothesis, H_A = Alternative Hypothesis, μ_A = Mean of AMOC sample, μ_S = Mean of SSO sample, and δ = non-inferiority/equivalence margin.

Study Design	Null Hypothesis	Alternative Hypothesis
Non-Inferiority	$H_0: \mu_A - \mu_S \leq -\delta$	$H_A: \mu_A - \mu_S > -\delta$
Equivalence	$H_0: \mu_A - \mu_S \leq -\delta$ or $\mu_A - \mu_S \geq \delta$	$H_A: \mu_A - \mu_S > -\delta$ and $\mu_A - \mu_S < \delta$
Superiority	$H_0: \mu_A - \mu_S = 0$	$H_A: \mu_A - \mu_S \neq 0$

difference nor equivalence. We submit that non-inferiority testing simplifies and clarifies the results for any comparison where we are looking for “equal to or better than”.

To demonstrate non-inferiority, a two-dimensional vertical grid framed by vertical lines at $\pm \delta$ (delta defined below) was created. The mean and confidence interval of the differences were computed. Non-inferiority is demonstrated when the mean and confidence interval are above the negative (–) delta. Non-inferiority and equivalence are demonstrated when the mean and confidence interval of the differences are between the negative (–) delta and positive (+) delta. Non-inferiority and superiority are demonstrated when the mean and confidence interval are above zero. It is possible to have results that are simultaneously equivalent, superior, and non-inferior. If the mean and confidence interval of the differences are above zero but below positive (+) delta, equivalence, superiority, and non-inferiority apply. If the mean and confidence interval of the differences cross the negative (–) delta line, the test is “non-inferiority not found” (Mascha and Sessler, 2011) (Fig. 1).

The confidence intervals can be set at 90%, 95%, or 97.5%. If testing at the 95% level, standard superiority testing can be read from the graphs generated during the analysis. A 95% confidence interval is used in this example as it is the regulatory recommendation in the medical field (Greene et al., 2008), and is most often used in current studies testing for non-inferiority. The mean and associated 95% confidence interval of the differences are then placed into the grid at the x and y coordinates of the mean with the confidence interval of each centered on the coordinates of the mean and aligned parallel to the x-axis.

The choice of the non-inferiority margin, denoted by delta, is derived from what is considered to be the largest clinically or operationally insignificant difference as determined by a group of experts in the area of interest. This means that in FRMS studies, the smallest operationally meaningful value is used as the delta margin (Piaggio et al., 2012; Walker and Nowacki, 2011). The FAA was consulted for maximum tolerable differences ($\pm \delta$) for the SPIs and the differences ($\pm \delta$) were set at ± 30 min for in-flight sleep, ± 0.31 responses/second for the PVT, and ± 1 for both the SP and KSS. These delta values should be regarded as provisional and may be subject to iterative improvement. To show how non-inferiority relates to equivalence and superiority, we have added blue reference arrows to our non-inferiority figures and included both delta margins.

3. Results

3.1. Flight duration

Flight Duration is the total flight time, measured from block out to block in. In the present study, the AMOC flight was scheduled to be longer than the SSO flight. This was confirmed by measurement, as shown in Fig. 2. The results for Flight Duration (hours), Total Inflight Sleep (hours), PVT mean speed at TOD, SP at TOD, and KSS at TOD have been included in Table 2.

3.2. Total in-flight sleep

The AMOC flight was scheduled to be longer than the SSO flight, therefore it was expected to allow for more in-flight sleep opportunity. This was confirmed by measurement, as shown in Figs. 3 and 4.

3.3. PVT mean speed as a function of flight type and flight direction

PVT mean speed was measured at TOD on the AMOC flight compared to the SSO flight. PVT mean speed showed non-inferior performance for the AMOC compared to the SSO, as shown in Figs. 5 and 6.

AMOC Compared to SSO

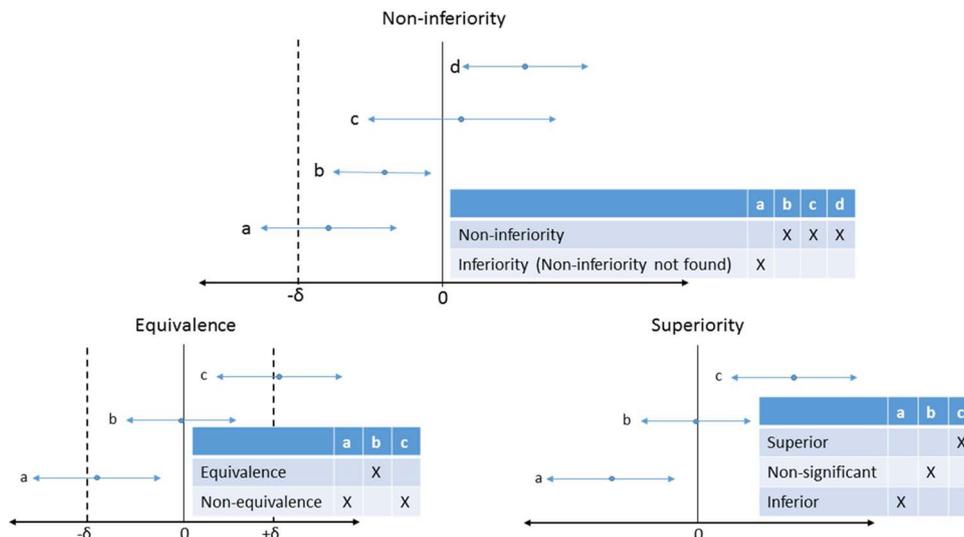


Fig. 1. Sample confidence intervals and interpretations of non-inferiority, superiority, and equivalence trials. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article).

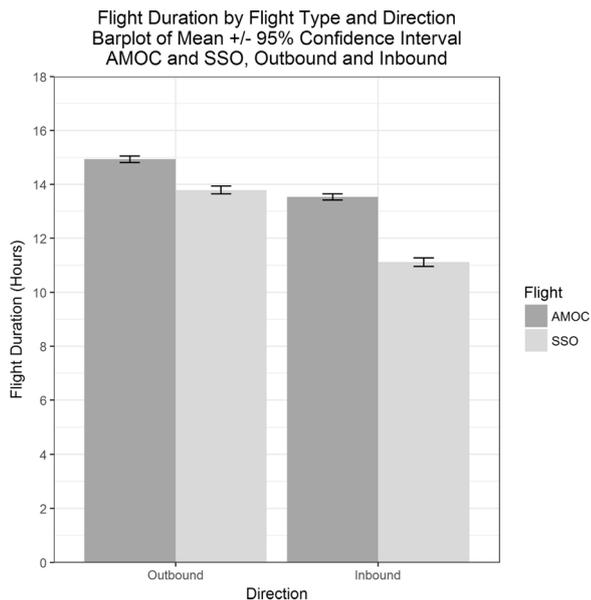


Fig. 2. Bar chart shows mean (arithmetic average) flight duration as a function of flight type (AMOC vs. SSO) and flight direction (outbound vs. inbound); error bars indicate the 95% confidence interval for the mean.

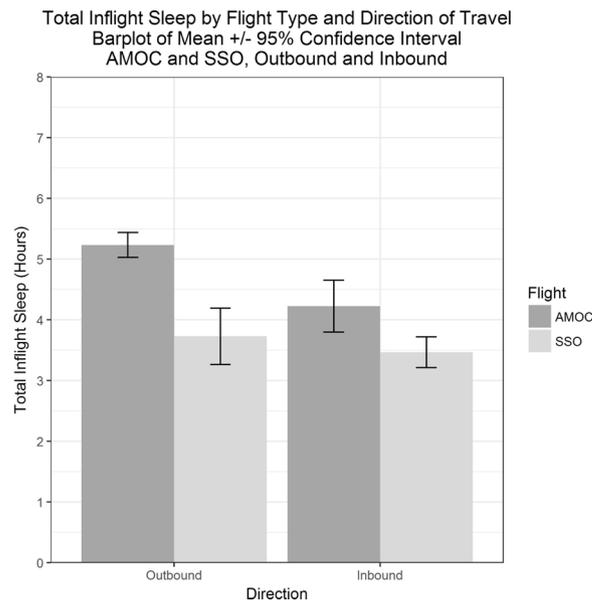


Fig. 3. Bar chart shows mean (arithmetic average) total inflight sleep as a function of flight type (AMOC vs. SSO) and flight direction (outbound vs. inbound); error bars indicate the 95% confidence interval for the mean.

3.4. Samn-Perelli Fatigue Scale as a function of flight type and flight direction

Fatigue was measured at TOD on the AMOC flight compared to the SSO flight. The Samn-Perelli Fatigue Scale showed non-inferior performance for the AMOC compared to the SSO, as shown in Figs. 7 and 8.

3.5. Karolinska Sleepiness Scale as a function of flight type and flight direction

Sleepiness was measured at TOD on the AMOC flight compared to the SSO flight. The Karolinska Sleepiness Scale showed non-inferior performance for the AMOC compared to the SSO, as shown in Figs. 9

Table 2

Result table recording mean ± 95% confidence interval of Flight Duration (hours), Total Inflight Sleep (hours), PVT mean speed at TOD, SP at TOD, and KSS at TOD.

	AMOC Outbound	SSO Outbound	AMOC Inbound	SSO Inbound
Flight Duration (hours)	14.9 ± 0.1	13.8 ± 0.1	13.5 ± 0.1	11.1 ± 0.2
Total Inflight Sleep (hours)	5.2 ± 0.2	3.7 ± 0.5	4.2 ± 0.4	3.5 ± 0.3
Performance (PVT mean speed) at TOD	3.60 ± 0.17	3.61 ± 0.16	3.63 ± 0.15	3.67 ± 0.20
SP at TOD	3.6 ± 0.5	3.6 ± 0.4	3.9 ± 0.4	3.6 ± 0.4
KSS at TOD	4.5 ± 0.7	4.4 ± 0.6	5.1 ± 0.7	4.5 ± 0.6

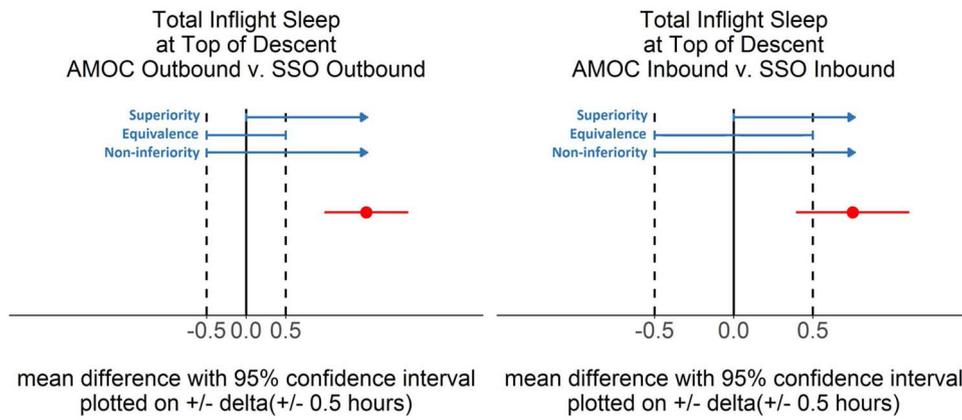


Fig. 4. Left: Non-inferiority plot shows non-inferior total inflight sleep (total inflight sleep is equal to or longer than) for the outbound AMOC vs. SSO. Right: Non-inferiority plot shows non-inferior total inflight sleep (total inflight sleep is equal to or longer than) for the inbound AMOC vs. SSO. The FAA was consulted for maximum tolerable differences (\pm delta).

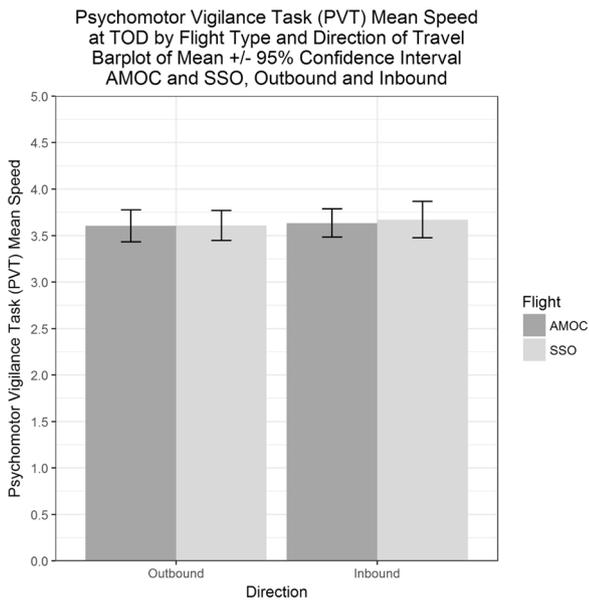


Fig. 5. Bar chart shows PVT mean speed at TOD as a function of flight type (AMOC vs. SSO) and flight direction (outbound vs. inbound); error bars indicate the 95% confidence interval for the mean.

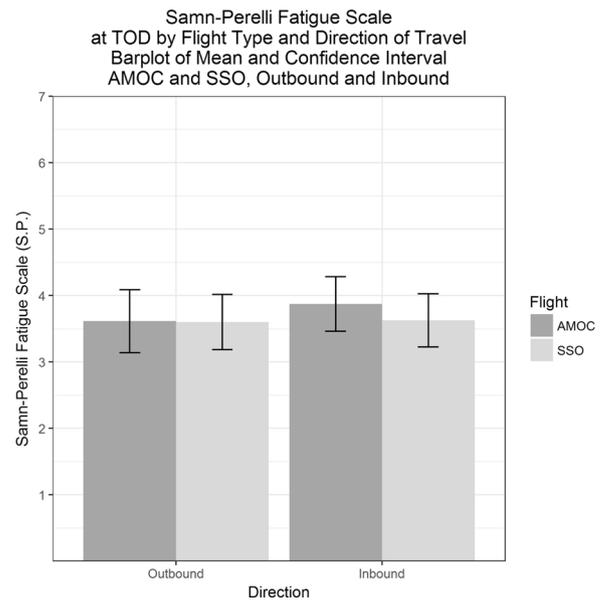


Fig. 7. Bar chart shows mean (arithmetic average) Samn-Perelli Fatigue Scale scores at TOD as a function of flight type (AMOC vs. SSO) and flight direction (outbound vs. inbound); error bars indicate the 95% confidence interval for the mean.

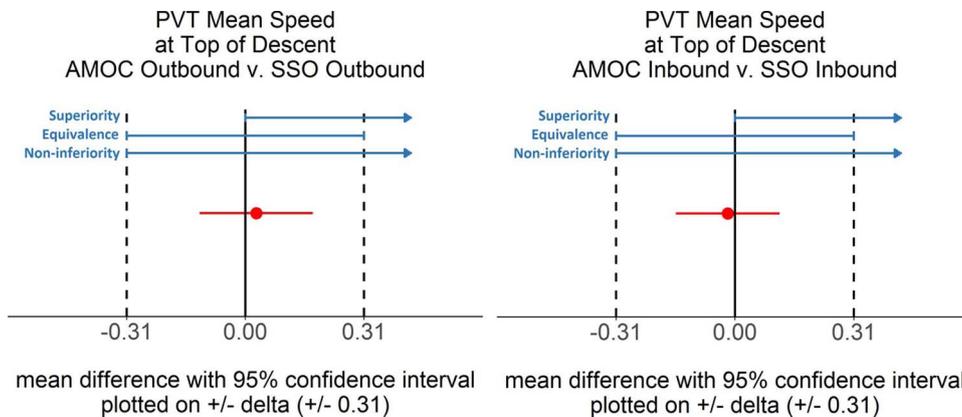


Fig. 6. Left: Non-inferiority plot shows non-inferior PVT mean speed at TOD (PVT mean speed is equal to or better than) for the AMOC vs. SSO. Right: Non-inferiority plot shows non-inferior PVT mean speed at TOD (PVT mean speed is equal to or better than) for the AMOC vs. SSO. The FAA was consulted for maximum tolerable differences (\pm delta).

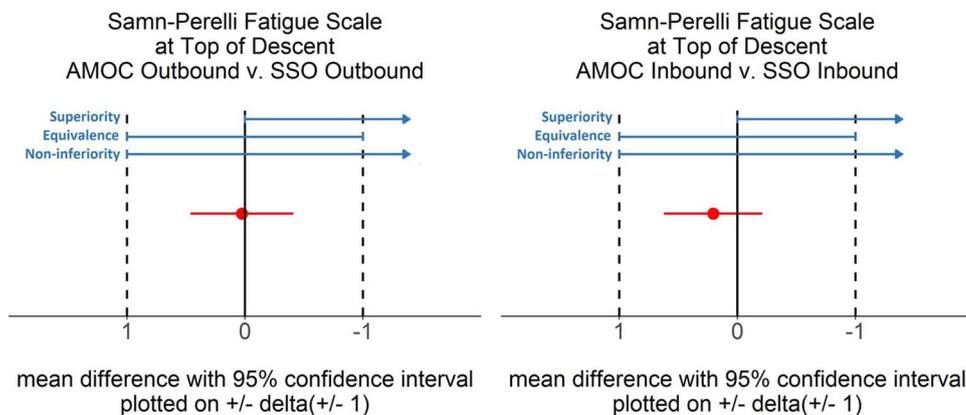


Fig. 8. Left: Non-inferiority plot shows non-inferior fatigue at TOD (fatigue is equal to or less than) for the outbound AMOC vs. SSO. Right: Non-inferiority plot shows non-inferior fatigue at TOD (fatigue is equal to or less than) for the inbound AMOC vs. SSO. The FAA was consulted for maximum tolerable differences (\pm delta). Note the x-axis and arrows are flipped with negative one to the right.

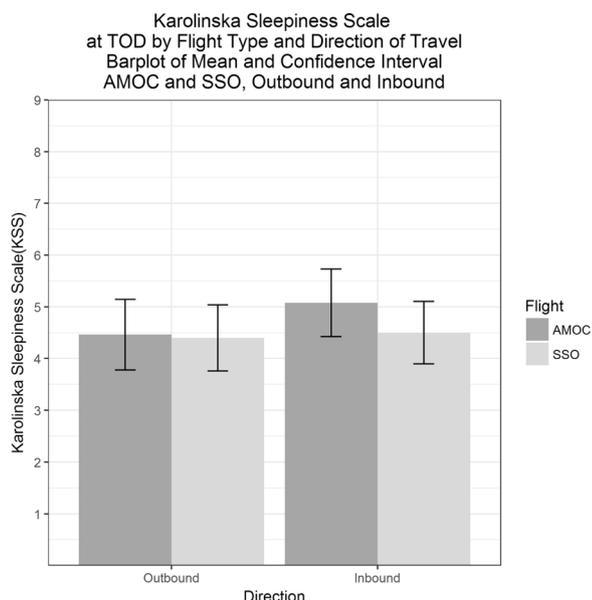


Fig. 9. Bar chart shows mean (arithmetic average) Karolinska Sleepiness Scale scores at TOD as a function of flight type (AMOC vs. SSO) and flight direction (outbound vs. inbound); error bars indicate the 95% confidence interval for the mean.

and 10.

4. Discussion and conclusions

The findings of the present study indicated that the US-Australia AMOC route was non-inferior (equivalent and/or superior) on all measures except inbound sleepiness. Though it was “non-inferiority not found” (Mascha and Sessler, 2011), the actual mean score on the

inbound KSS corresponded to “neither alert nor sleepy” and was deemed to be within the margins of safety to sustain operational performance. Given that seven of eight measures were non-inferior, and the eighth was indeterminate yet within the margins of safety to sustain operational performance, the AMOC flight was determined to be safe. In the event of a more decidedly mixed case, e.g., more than one SPI was indeterminate or inferior, a qualitative assessment could be made by evaluating comments from pilots with discrepant SPIs to understand if there were special circumstances at play. If after a qualitative assessment the situation remained unresolved, additional fatigue mitigations could be introduced and the route could be restudied. It is worth noting that with more than 220 pilots studied in the paradigm described, we have yet to find such a decidedly mixed case as imagined above.

This paper is the first to provide a detailed illustration of the non-inferiority method embedded in the context of a commercial aviation FRMS. Non-inferiority testing provides an elegant method to demonstrate in a clear, visual display how a non-compliant flight flown under an FRMS demonstrates at least an equivalent level of safety to a compliant flight – as required by FAA regulations. Non-inferiority is a superior analytical tool for comparing AMOC and SSO flights as it allows for a clear, interpretable explanation of the flight as being non-inferior; namely showing with clear graphics whether the AMOC flight is as safe as or safer than the SSO.

Non-inferiority and equivalence both require an a priori estimate of delta. There is no standardized, quantitative method for estimating delta. In the future it may be possible to set delta based on available data. A limitation of both equivalence testing and non-inferiority testing is that the deltas are presently set through the consensus of subject matter experts with some contribution from the limited data available. For example, the current delta values of ± 1 for SP and KSS may not indicate precise marginal differences in safety as fatigue and sleepiness increase. Thus, a systematic procedure to develop and periodically re-evaluate deltas for equivalence and non-inferiority testing is

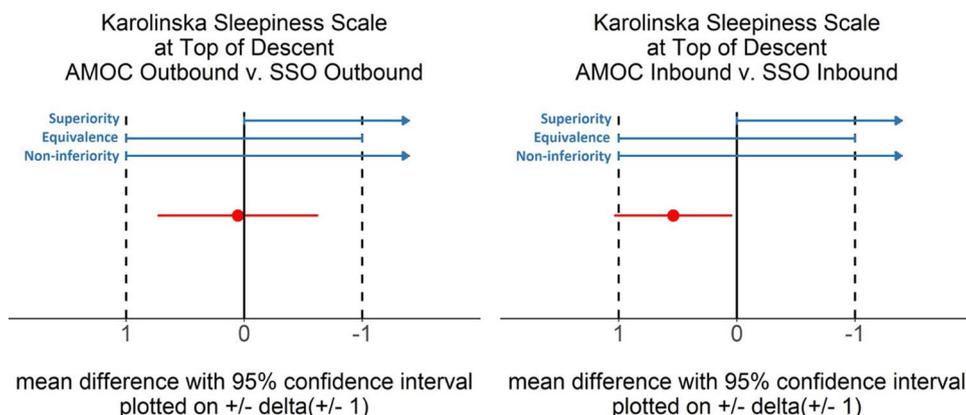


Fig. 10. Left: Non-inferiority plot shows non-inferior sleepiness at TOD (sleepiness is equal to or less than) for the outbound AMOC vs. SSO. Right: Non-inferiority plot shows non-inferiority not found (Mascha and Sessler, 2011) results for sleepiness at TOD (sleepiness is equal to or less than) for the inbound AMOC vs. SSO. The FAA was consulted for maximum tolerable differences (\pm delta). Note the x-axis and arrows are flipped with negative one to the right.

needed. We should work to ensure that deltas have been set through a well-documented approach driven by data, scientific consultation, and subject matter expertise, with the same deltas used throughout the aviation industry in the development and implementation of FRMSs.

This work extends the current body of literature on fatigue risk management by introducing non-inferiority testing as a clear, concise, and practical alternative to the analyses currently used in FRMSs in the United States and elsewhere. The non-inferiority model translates directly to any 24/7 operation to ensure safety. It has potential as the basis of a commercially viable tool in support of FRMSs. This comprehensive and easily interpretable non-inferiority analysis should be used in future FAA-mandated FRMS studies and in any analysis using a comparison to determine operational risk, taking the place of separate equivalence and superiority testing. Non-inferiority analysis enables the reader to see at a glance whether the data support the AMOC as being as safe as or safer than the SSO.

Funding Source

United Airlines provided financial support to conduct the research and prepare this article. United Airlines had no role in the study design, nor in collection, analysis and interpretation of the data, nor in the writing of the report, nor in the decision of submitting the paper for publication.

Conflicts of Interest

Conflicts of interest: none.

Acknowledgements

We would like to thank United Airlines for funding the studies, the Air Line Pilots Association, International (ALPA) representatives for supporting the studies, and the pilots that participated in the studies that supplied data for this paper. We would also like to thank Chip Benton and Ellen Birmingham at United Airlines for their support of the studies and this paper along with Ellen Birmingham's edits. Additionally, we would like to thank Rachelle Brown for her editing and proofreading.

References

Achermann, P., Borbely, A., 2003. Mathematical Models of Sleep Regulation. Retrieved

- from. <https://www.bioscience.org/2003/v8/s/1064/fulltext.htm>.
- Åkerstedt, T., Gillberg, M., 1990. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.* 52 (1–2), 29–37.
- Ancoli-Israel, S., Martin, J., Blackwell, T., Buenaer, L., Liu, L., Meltzer, L., Sadeh, A., Spira, A., Taylor, D., 2015. The SPSM guide to actigraphy monitoring: clinical and research applications. *Behav. Sleep Med.* 13 (Suppl.1), S4–S38.
- Belenky, G., Wu, L.J., Jackson, M.L., 2011. Occupational sleep medicine: practice and promise. *Prog. Brain Res.* 190, 189–203.
- Borbély, A., 1982. A two process model of sleep regulation. *Hum. Neurobiol.* 1, 195–204. http://dx.doi.org/10.1007/978-3-540-29678-2_6166.
- Borbély, A., Daan, S., Wirz-Justice, A., Deboer, T., 2016. The two-process model of sleep regulation: a reappraisal. *J. Sleep Regul.* 25, 131–143.
- Federal Aviation Administration (FAA), 2013. Fatigue Risk Management Systems (FAA AC 120-103A). Retrieved from. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-103A.pdf.
- Flight And Duty Limitations And Rest Requirements: Flightcrew Members, 14 C.F.R. § 117 (2012).
- Gander, P.H., Mulrine, H.M., Van den Berg, M.J., Smith, A.A., Signal, T.L., Wu, L.J., Belenky, G., 2013a. Effects of Extended Flight and Duty Durations on Flight Crew Fatigue. "Manuscript in preparation".
- Gander, P.H., Signal, T.L., Berg, M.J., Mulrine, H.M., Jay, S.M., Jim Mangie, C., 2013b. In-flight sleep, pilot fatigue and psychomotor vigilance task performance on ultra-long range versus long range flights. *J. Sleep Res.* 22 (6), 697–706.
- Greene, C.J., Morland, L.A., Durkalski, V.L., Frueh, B.C., 2008. Noninferiority and equivalence designs: issues and implications for mental health research. *J. Trauma. Stress* 21 (5), 433–439.
- Hursh, S.R., Balkin, T.J., Miller, J.C., Eddy, D.R., 2004. The Fatigue Avoidance Scheduling Tool: Modeling to Minimize the Effects of Fatigue on Cognitive Performance. International Civil Aviation Organization (ICAO), 2008. Annex 6 Operation of Aircraft 115. pp. 70–90.
- Lamond, N., Dawson, D., Roach, G.D., 2005. Fatigue assessment in the field: validation of a hand-held electronic psychomotor vigilance task. *Aviat. Space Environ. Med.* 76 (5), 486–489.
- Mascha, E.J., Sessler, D.I., 2011. Equivalence and Noninferiority Testing in Regression Models and Repeated-Measures Designs. Retrieved from. <https://insights.ovid.com/pubmed?pmid=21304155>.
- Piaggio, G., Elbourne, D.R., Pocock, S.J., Evans, S.J., Altman, D.G., Group, C., 2012. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 308 (24), 2594–2604.
- Rangan, S., Bowman, J.L., Hauser, W.J., McDonald, W.M., Lewis, R.A., Van Dongen, H.P.A., 2013. Integrated fatigue modeling in crew rostering and operations. *Can. Aeronaut. Space J.* 59 (1), 1–6.
- Samn, S.W., Perelli, L.P., 1982. Estimating Aircrew Fatigue: a Technique With Application to Airlift Operations. Retrieved from. https://www.researchgate.net/profile/Sherwood_Samn/publication/235150215_Estimating_Aircrew_Fatigue_A_Technique_with_Application_to_Airlift_Operations/links/0046352fce1ebe7bf3000000.pdf.
- Thorne, D.R., Johnson, D.E., Redmond, D.P., Sing, H.C., Belenky, G., Shapiro, J.M., 2005. The Walter Reed palm-held psychomotor vigilance test. *Behav. Res. Methods* 37 (1), 111–118.
- Walker, E., Nowacki, A.S., 2011. Understanding equivalence and noninferiority testing. *J. Gen. Intern. Med.* 26 (2) 192-1.