Quantifying the Safety Benefits of a Digital Copilot in General Aviation

Kevin J. Burns, Craig A. Bonaceto, Steven L. Estes, John R. Helleburg
The MITRE Corporation, 7515 Colshire Drive, McLean, VA, 22102, USA

Digital Copilot is a prototype information system providing cognitive assistance to pilots in general aviation. This paper presents methods and results of an analysis quantifying the expected safety benefits of the system. The methods include: Cognitive Performance Analysis, to quantify pilot workload in flight activities; Human Reliability Analysis, to quantify pilot error probabilities; and Probabilistic Risk Assessment, to quantify accident sequence frequencies. These methods are used to model flight risk with and without pilots' use of Digital Copilot, and the comparison provides a quantitative measure of expected safety benefits assuming the prototype system is adopted by pilots. The results suggest that Digital Copilot technology has the potential to significantly reduce the rates of total and fatal accidents in general aviation. The same methods also enable a detailed analysis of individual pilot actions and system features, to quantify their risk importance and identify the most promising opportunities for further system improvements.

I. Introduction

APROXIMATELY 75% of accidents in general aviation are attributed to pilot error [1-5], as a pilot must perform numerous information processing tasks while at the same time controlling the aircraft through all phases of flight. These tasks include: scanning for traffic; accessing charts, checklists, manuals, and weather reports; planning and re-planning en route as required by weather conditions and fuel supply; and communicating with air traffic controllers. The complexity of these tasks is compounded in flights operated by a single pilot. Recent efforts to improve the safety of general aviation have attempted to address pilot workload by applying concepts of crew resource management [6] from commercial aviation, where pilots fly with trained copilots and each individual has assigned roles.

In one such effort, a prototype mobile computing application known as Digital Copilot [7] has been developed to demonstrate technology for providing cognitive assistance to pilots. Using a speech recognition interface, the pilot can receive answers to questions and request information to be read aloud. For example, when asked "Will the tower be open?" (at the destination airport), Digital Copilot responds after estimating the arrival time and accessing the tower schedule stored in its records. The system is also able to anticipate a pilot's information needs, such as the ATIS (Automatic Terminal Information Service) frequency, by inferring the phase of flight and expected arrival airport.

This prototype system has been designed to reduce pilot workload, and thereby prevent errors that may lead to accidents. The question is: *To what extent does the Digital Copilot reduce pilot workload, and to what extent do these workload reductions prevent pilot errors and avert accidents?* This question cannot be answered in retrospective fashion, because historical data do not include operations with pilots' use of Digital Copilot. Therefore, our approach uses prospective methods of Probabilistic Risk Assessment, Human Reliability Analysis, and Cognitive Performance Analysis to quantify the safety of general aviation both with and without pilots' use of Digital Copilot.

A previous study [8] used standard methods of Human Reliability Analysis and Probabilistic Risk Assessment to quantify human errors and accident sequences with and without pilots' use of Digital Copilot. However, that study was limited by existing methods for computing Human Error Probabilities as a function of various Performance Shaping Factors, including Stress/Stressors, Complexity, Available Time, and Ergonomics/Human-Machine Interface. The treatment of Ergonomics/Human-Machine Interface was especially problematic, as existing methods do not identify Workload as a Performance Shaping Factor or quantify the expected impacts of Workload on Human Error Probabilities. The present study overcomes these limitations by defining a new Performance Shaping Factor for Workload, and by developing a novel Workload-Reliability Correlation that quantifies Human Error Probabilities as a function of Workload.

These advancements in Cognitive Performance Analysis, detailed in Section II, enable more accurate and credible predictions of the expected safety benefits provided by Digital Copilot technology, discussed in Section III.

II. Methods

Three classes of methods are used in this study, namely: Probabilistic Risk Assessment (PRA); Human Reliability Analysis (HRA); and Cognitive Performance Analysis (CPA). The key inputs, outputs, and use cases for these methods are summarized in Table 1, with details of each method discussed in subsections below.

Table 1 Input, output, and use case for each method

Method	Input	Output	Use Case
PRA	Human Error Probabilities	Accident Sequence Frequencies	With and Without Digital Copilot
HRA	Performance Shaping Factors	Human Error Probabilities	Without Digital Copilot
CPA	Subjective Workload Ratings	Human Error Probabilities	With Digital Copilot

A. Probabilistic Risk Assessment

PRA [9] is a prospective approach to quantifying the risks of human-system operations under hazardous conditions, using event trees to identify accident sequences that stem from failures of critical safety functions, and using fault trees [10] to identify combinations of basic events that lead to failures of critical safety functions. Each basic event is either a human error, system failure, or world event such as the presence of a hazard, and each basic event is quantified with a probability of occurrence.

The present study is focused on pilot-related accidents that might be mitigated by features of Digital Copilot technology. Therefore, all basic events in our PRA fault trees represent human errors or world events that set the context for human errors, and all accident sequences in our PRA event trees represent combinations of these basic events. A total of eight event trees are used to model sequential phases of flight as follows: Taxi-out, Takeoff, Climb, En route, Descent, Approach, Landing, and Taxi-in. A total of 30 fault trees are used to model failures of critical safety functions appearing in the event trees. A total of 12 world events (see Table 2) and 38 pilot actions (see Table 3) appear as fault tree basic events. The trees are linked and solved using standard methods [11] to compute event tree accident sequences as combinations of fault tree basic events, where each accident sequence represents a unique combination of human errors and world events. Similar sequences are then binned to obtain subtotals for 14 accident sequence bins, and the bins are summed to obtain an overall accident sequence frequency.

As outlined above, PRA was performed for two cases that differed in the underlying probabilities of human errors used as inputs to the analysis (see Table 1). First, a baseline PRA model was quantified with Human Error Probabilities obtained from HRA, assuming pilots were not using the Digital Copilot. Then, a revised PRA model was quantified with Human Error Probabilities obtained from CPA, assuming pilots were using the Digital Copilot. Results of the two models were then compared to compute the numbers of accidents expected to be averted by pilots' use of the Digital Copilot, as discussed in Section III.

Table 2 World event probabilities

Phase of Flight	World Event	Probability
Taxi-out	Collision hazard present	0.1
Takeoff	Collision hazard present	0.1
Climb	Collision hazard present	0.01
En route	Collision hazard present	0.00001
Descent	Collision hazard present	0.001
Approach	Collision hazard present	0.1
Landing	Collision hazard present	0.1
Taxi-in	Collision hazard present	0.1
Takeoff	Departure airport controlled	0.40
Approach	Arrival airport controlled	0.40
En route	Hazardous weather en route	0.05
En route	Hazardous weather at destination	0.05

Table 3 Pilot actions, types, and Performance Shaping Factor multipliers for Stress/Stressors (s), Complexity (c), and Available Time (t)

Phase	Pilot Action	Type	s	c	t	s*c*t
Taxi-out	1. Fail to hold short as required	K	1	1	1	1
Taxi-out	2. Fail to follow proper taxi route to proper runway	R	1	2	1	2
Taxi-out	3. Fail to avoid collision with hazard	K	5	1	10	50
Taxi-out	4. Fail to manage systems and maintain control	S	1	1	1	1
Takeoff	5. Fail to obtain clearance for takeoff (controlled airport)	R	1	1	1	1
Takeoff	6. Fail to verify runway clear of hazards (non-controlled airport)	K	1	1	1	1
Takeoff	7. Fail to avoid collision with hazard	K	5	1	10	50
Takeoff	8. Fail to manage systems and maintain control (wake turbulence)	R	2	2	1	4
Takeoff	9. Fail to ensure sufficient runway remaining	K	2	2	1	4
Takeoff	10. Fail to reject takeoff if insufficient runway remaining	K	5	1	1	5
Climb	11. Fail to maintain proper climb profile or heading	R	1	1	1	1
Climb	12. Fail to avoid collision with hazard	K	5	1	10	50
Climb	13. Fail to manage systems and maintain control	S	1	1	1	1
En route	14. Fail to conform to flight plan / airspace requirements	R	1	2	1	2
En route	15. Fail to avoid collision with hazard	K	5	1	10	50
En route	16. Fail to avoid hazardous weather en route	K	2	2	1	4
En route	17. Fail to avoid loss of control in hazardous weather en route	K	5	1	10	50
En route	18. Fail to manage systems and maintain control, fuel sufficiency	R	1	2	1	2
En route	19. Fail to avoid hazardous weather at destination	K	2	2	1	4
En route	20. Fail to avoid loss of control in hazardous weather at destination	K	5	1	10	50
Descent	21. Fail to maintain proper descent profile or heading	R	1	1	1	1
Descent	22. Fail to avoid collision with hazard	K	5	1	10	50
Descent	23. Fail to manage systems and maintain control	S	1	1	1	1
Approach	24. Fail to conform to ATC directives (controlled airport)	R	1	1	1	1
Approach	25. Fail to ID proper runway / approach (non-controlled airport)	K	1	2	1	2
Approach	26. Fail to avoid collision with hazard	K	5	1	10	50
Approach	27. Fail to detect infeasible approach and execute go around	K	5	1	1	5
Approach	28. Fail to manage systems and maintain control (wake turbulence)	R	2	2	1	4
Landing	29. Fail to obtain clearance for landing (controlled airport)	R	1	1	1	1
Landing	30. Fail to verify runway clear of hazards (non-controlled airport)	K	1	1	1	1
Landing	31. Fail to avoid collision with hazard	K	5	1	10	50
Landing	32. Fail to manage systems and maintain control	S	2	2	10	40
Landing	33. Fail to ensure sufficient runway remaining	K	2	2	1	4
Landing	34. Fail to go around if insufficient runway remaining	K	5	1	1	5
Taxi-in	35. Fail to hold short as required	K	1	1	1	1
Taxi-in	36. Fail to follow proper taxi route from proper runway	R	1	2	1	2
Taxi-in	37. Fail to avoid collision with hazard	K	5	1	10	50
Taxi-in	38. Fail to manage systems and maintain control	S	1	1	1	1

B. Human Reliability Analysis

HRA [12] is a prospective approach to quantifying Human Error Probabilities (HEPs), based on various Performance Shaping Factors (PSFs) that account for the situational context in which human actions are accomplished. These PSFs are modeled as multipliers to nominal HEPs, which in turn model fundamentally different types of human actions. In our approach, each of the 38 pilot actions appearing as a fault tree basic event was classified as one of three types, in accordance with the qualitative cognitive performance model proposed by Rasmussen [13]. These three types are Knowledge-based (K), Rule-based (R), and Skill-based (S), where: K refers to goal-directed behavior involving reasoning and planning to diagnose and act; R refers to procedural behavior in accordance with explicit, step-by-step guidance; and S refers to sensory-motor behavior accomplished without conscious control. The 38 actions and their types (K, R, S) are listed in Table 3, along with PSF multipliers discussed further below.

The numerical values of nominal HEPs are assumed to be ordered such that K >> R > S, consistent with the relative complexity of the three types. However, unlike the standard method of [12], we do not assume numerical values for nominal error probabilities. Instead, we compute nominal values for K, R, and S using historical accident rate data aligned with our PRA accident sequences, as discussed in Section III. This is because the standard method was developed for analysis of accidents in nuclear power operations, and general aviation differs in the following respects:

First, general aviation usually involves a single pilot flying without a copilot and executing highly-practiced actions, as opposed to the multi-person crews of control room and auxiliary operators responding to rarely-occurring nuclear power plant accidents. Second, timeframes for pilot actions in general aviation are typically seconds or several minutes, as opposed to the many minutes or hours that are often involved in responding to nuclear accidents. Finally, unlike nuclear accidents that rarely occur, large amounts of historical data from aviation safety databases are available to support estimation of pilot error probabilities in flight operations.

Along with nominal HEP values discussed above, the standard method of HRA [12] uses numerical PSF multipliers to modify the nominal HEPs and thereby account for the performance shaping context of different situations in which actions of types K, R, or S are accomplished. The PSFs are generic factors, including multipliers for Stress/Stressors, Complexity, and Available Time that are expected to apply in any domain, unlike the nominal HEPs for K, R, and S that are expected to vary from domain to domain. Therefore, our approach uses the PSFs and associated multipliers of the standard method [12], with only minor modifications described in [8] to ensure that PSF multipliers for Stress/Stressors, Complexity, and Available Time are applied consistently across actions of type K, R, and S. The resulting set of PSF levels and associated multiplier values, derived from the standard method [12], are as follows:

- Stress/Stressors (s): Nominal (1), High (2), Extreme (5)
- Complexity (c): Low (0.1), Nominal (1), Moderate (2), High (5)
- Available Time (t): Extra Time (0.1), Nominal (1), Barely Adequate Time (10)

In applying the standard method of HRA [12], one or more Subject Matter Experts (SMEs) are provided with a detailed description of each action and asked to rate the associated level of each PSF (s, c, t). Although it is common practice in HRA to obtain PSF judgments from only one or two SMEs, we obtained consensus ratings from four SMEs, including two general aviation experts who are also human factors engineers, and two risk analysts who are HRA specialists. The resulting PSF multipliers (s, c, t) for all 38 pilot actions in the PRA model are listed in Table 3. The corresponding HEP for an action of type K, or R, or S is computed as K*s*c*t, or R*s*c*t, or S*s*c*t, respectively.

Besides nominal HEPs and PSF multipliers, the PRA model also requires values for the probabilities of various world events listed in Table 2. Probabilities for controlled/non-controlled airports were obtained from statistics on airport operations, and probabilities for all other world events were obtained as judgments from the same SMEs that provided judgments of PSFs. These judgments are supported by empirical benchmarking of the integrated PRA model, which includes nominal HEPs, PSFs, and world event probabilities. This benchmarking, described in Section III, compares the computed accident sequence frequencies to historical accident frequencies for pilot-related accidents in general aviation [1-5].

C. Cognitive Performance Analysis

As described above, the baseline model (without pilots' use of Digital Copilot) uses three PSFs (s, c, t) to model how nominal HEPs (K, R, S) are affected by the performance-shaping context of Stress/Stressors, Complexity, and Available Time. In a previous study [8], a fourth PSF for Ergonomics/Human-Machine Interface (HMI) was then used in accordance with the standard method of HRA [12] to quantify a revised model that included pilots' use of Digital Copilot. However, that approach suffered from several important limitations, as follows:

First, the benefits of Digital Copilot were modeled using a PSF for Ergonomics/HMI that could not be benchmarked empirically, because the historical data do not include operations with Digital Copilot. Second, the three PSFs that were benchmarked in the baseline model, namely Stress/Stressors, Complexity, and Available Time, would all be affected by Ergonomics/HMI. Ideally a PSF for Ergonomics/MHI would be directly related to these three PSFs, and thereby enable more integrated analysis of safety with and without pilots' use of Digital Copilot. Finally, in accordance with the standard method of HRA [12], the PSF multiplier for Ergonomics/HMI was limited to a numerical value of either 1.0 (Nominal) or 0.5 (Good) based on SME judgments about the effectiveness of the HMI. Ideally the multiplier would instead vary along a continuum, and thereby provide more fine-grained modeling of how nominal HEPs are affected by the Ergonomics/HMI of a system such as Digital Copilot.

The present study overcomes these limitations by replacing the PSF for Ergonomics/HMI with a new PSF for Workload, and by measuring Workload on a continuous scale aligned with the three categorical PSFs of Stress/Stressors, Complexity, and Available Time. This enables benchmarking of the new PSF for Workload against the baseline PSFs of Stress/Stressors, Complexity, and Available Time, which in turn have been benchmarked against historical accident rate data using the integrated PRA model. The approach includes an instrument for measuring Workload and a Workload-Reliability Correlation, as discussed in two subsections below.

1. Instrument for Measuring Workload

Use of a new PSF for Workload requires that SMEs provide ratings of Workload for pilot actions in the PRA model, both with and without pilots' use of the Digital Copilot. A number of instruments exist for obtaining such ratings of Workload, including the NASA Task Load Index (TLX) [14], the Subjective Workload Assessment Technique (SWAT) [15], and the Overall Workload (OW) [16] scale. NASA-TLX requires that SMEs provide ratings on a scale of 0-20 along six different dimensions of Workload. Three of these dimensions are directly related to our three PSFs of Stress/Stressors, Complexity, and Available Time, namely Frustration Level, Mental Demand, and Temporal Demand. Two other dimensions, referred to as Mental Effort and Performance, appear to be somewhat redundant, and another dimension, referred to as Physical Demand, is not relevant to our study of cognitive workload. SWAT includes only three dimensions, namely Psychological Stress, Mental Effort, and Time Load, which are aligned with our PSFs of Stress/Stressors, Complexity, and Available Time. OW obtains a direct measure of Overall Workload, without the need to combine measures made along individual dimensions as in NASA-TLX and SWAT.

Our instrument includes elements of NASA-TLX, SWAT, and OW as follows: First, we define Stress Level, Mental Demand, and Temporal Demand as three factors to be considered in ratings of Overall Workload. The definitions of these factors are taken directly from NASA-TLX, and the three factors are also consistent with the three dimensions of SWAT. Then, similar to OW, we obtain integer ratings of Overall Workload on a scale of 0-20 like that used in NASA-TLX. When providing their ratings, SMEs are instructed that Overall Workload is to be rated in accordance with the following definition: *Considering Stress Level, Mental Demand, and Temporal Demand, how much overall load does this task impose to achieve a sufficient level of performance? What is the overall difficulty of this task? Could you perform other tasks at the same time, or does this task demand your full attention?*

Using the scale of 0-20, a rating of 10 represents a nominal level of Workload. As such, the instrument elicits ratings on a ten-point scale of 0-9 lower than nominal Workload and a ten-point scale 11-20 higher than nominal Workload. When providing their ratings, SMEs are instructed to use these scale ranges below and above nominal, as they see fit, to characterize levels of Overall Workload that are judged to be lower, higher, or equal to the nominal anchor of 10.

This instrument was implemented in the form of a spreadsheet describing each of the 38 pilot actions in our PRA model, along with a dropdown menu for rating Overall Workload on the scale of 0-20. For each action, the spreadsheet also described relevant features of the Digital Copilot that can potentially affect the action. Using the spreadsheet, a total of 26 general aviation pilots each provided ratings of expected Overall Workload for all 38 pilot actions assuming flight without Digital Copilot. Then, for the 19 actions potentially affected by Digit Copilot, each pilot provided ratings of Overall Workload assuming flight with Digital Copilot. The participants were recruited with a "call for pilots" that described the purpose and method of our study, sent via email to a local (Washington, D.C., area) listing of general aviation pilots. As a requirement for participation, all pilots reviewed a detailed training briefing describing features of Digital Copilot. The training also included instructions on the method and scale to be used for reporting judgments of Workload on the spreadsheet. After this training, completion of the spreadsheet by a pilot required approximately one hour. All materials were distributed to participants via email, and the spreadsheets with ratings were collected via email as they were completed at the convenience of each individual pilot over a two-week period. All 26 pilots were Part 91 certificated, and none of the pilots was employed by The MITRE Corporation or any government agency. Of the 26 pilots: 17 (65%) were private pilots with instrument rating; 7 (27%) were commercial pilots with instrument rating; and 2 (8%) were airline transport pilots. The pilots' logged flight hours ranged from 200 to 3200, with a mean of 1228 and standard deviation of 861.

Each pilot provided ratings first assuming flight using a familiar Electronic Flight Bag (EFB) without Digital Copilot technology, and then assuming flight using Digital Copilot technology integrated into the same EFB. This was to ensure that changes in Workload with Digital Copilot would be attributed to the new technology and not existing technology available in today's EFBs. All except one pilot used an EFB on a regular basis, and the most popular EFB was Foreflight, which was used by 17 (65%) of the 26 pilots. When providing their ratings, pilots were instructed to assume the following situational conditions:

- A daytime flight under Visual Flight Rules with clear weather conditions and average winds, except for specific actions that are identified in the spreadsheet as being associated with hazardous weather conditions.
- A somewhat familiar airport with two runways and moderately complex taxi routes.
- A fixed-wing, single-engine, propeller driven aircraft, such as a Cessna 172, with a standard "six-pack" set of primary flight instruments.
- A familiar tablet-based EFB application, such as ForeFlight, Garmin Pilot, or WingX, for the case of flight without Digital Copilot. The same EFB, but with Digital Copilot technology integrated into the application, for the case of flight with Digital Copilot.

The resulting ratings of expected Workload, and associated impacts on accident sequence frequencies, are discussed in Section III.

2. Workload-Reliability Correlation (WORC)

Numerous studies have measured subjective Workload in various tasks of aviation and other domains, using instruments similar to ours described above. However, we are not aware of any existing method for converting measured Workload quantities to human error probabilities, as needed for quantifying fault tree basic events and computing event tree accident sequences. Here we present a novel method for doing so in the form of a Workload-Reliability Correlation (WORC), which is a key component of our analytic approach to quantifying the safety benefits of Digital Copilot.

Mathematically, WORC uses probability P to model how a human's propensity for failure on a task increases with Workload W. The underlying assumption is that human reliability is bounded by a finite capability for managing Workload, such that P increases with W in a manner that depends on P itself. More specifically: the marginal increase in failure probability dP/dW at a given level of Workload is assumed to be proportional to the value of P at that level of Workload; and this increase occurs at a probability 1-P, which represents the probability that failure has not already occurred at the current level of Workload. Taken together, these two effects lead to the following differential equation:

$$dP/dW = \omega * P * (1-P),$$

where ω is a constant of proportionality. Integrating this equation yields a logistic function as follows:

$$P = 1 / (1 + e^{\omega*(X - W)}),$$

where X is the value of Workload at which P = 0.5.

This theoretical model is supported by empirical data from laboratory experiments, which measure reliability (1-P) in the range 0-1 as a function of task difficulty and find that human performance exhibits the characteristic S-shape of a logistic function [17]. Elsewhere in human performance modeling, a logistic function also appears in the well-known Rasch [18] equation $P = 1 / (1 + e^{a-d})$ for the probability (P) of incorrect item response as a function of ability (a) and task difficulty (d).

For pilot actions in general aviation, human error probabilities are orders of magnitude less than 0.5, and when P is less than 0.5 the logistic function is very closely approximated by a simple exponential function. Therefore, we can reformulate the WORC equation as follows:

$$P = P_0 * e^{ML},$$

where L is a measure of Workload defined as L = (W-10)/5, and W is Workload measured on a scale of 0-20 in accordance with the instrument described above. This rescaling is performed so that $P = P_0$ is the nominal HEP (K, R, or S) corresponding to the nominal level of Overall Workload W = 10 where L = 0 and $e^{ML} = 1$. The resulting formulation of WORC can now be related directly to the PSF approach used in our benchmarked HRA, by mapping vectors of categorical PSF multipliers [Low, Nominal, High, Extreme] to associated values in the range of the numerical Workload scale [0, 10, 15, 20]. The value of M in the WORC equation e^{ML} is then adjusted to fit the exponential function to corresponding values of the total PSF multiplier (s*c*t) in the baseline model, as shown in Fig. 1. Note that the plot of Fig. 1 contains only six data points, because SME judgments of PSFs in the baseline model resulted in only six unique combinations of the PSF multipliers for Stress/Stressors, Complexity, and Available Time (see Table 3). As illustrated in Fig. 1, the benchmarked WORC equation e^{ML} with a value of M = 2.8 provides an excellent fit to values of the combined PSF multipliers (s*c*t).

As outlined above, WORC computes a multiplier (e^{ML}) that predicts how Workload L affects a pilot error probability P. The same correlation enables calculation of another multiplier, denoted w = P'/P = $e^{ML'}/e^{ML}$, which predicts how a change in Workload from L to L' will change the pilot error probability from P to P'. Substituting the expressions L' = (W'-10)/5 and L = (W-10)/5, and using the value M = 2.8, yields the following expression for w as a function of ΔW: w = $e^{-0.56*\Delta W}$, where ΔW = W – W'. This Workload multiplier w, computed from ΔW for a given pilot action with baseline error probability P, is then used to obtain a revised error probability P' = w*P that accounts for the change in Workload from W to W'.

The values of w, computed in this manner, are provided in Table 4 for the 19 pilot actions in Table 3 that are affected by Digital Copilot. These values of w, and the underlying values of ΔW obtained from pilots' ratings of expected Workload, are discussed further in Section III.

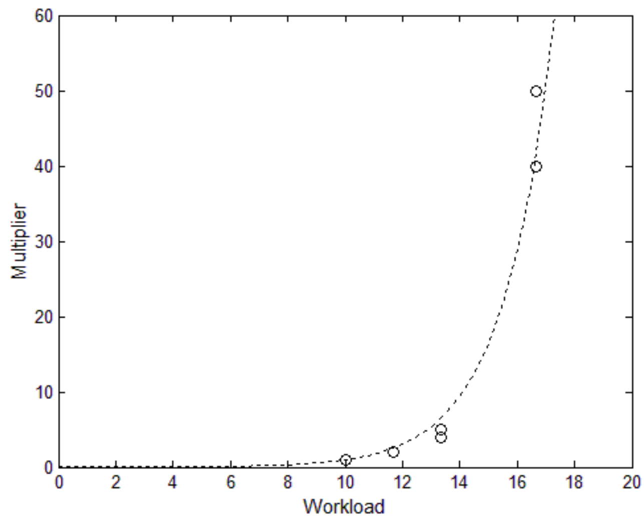

Fig. 1 Fit of Workload-Reliability Correlation (e^{ML}, with M = 2.8) to combined Performance Shaping Factor multipliers (s*c*t)

Table 4 Pilot actions, system features, and Performance Shaping Factor multiplier (w) for Workload affected by Digital Copilot

Pilot Action (Phase) System Feature	**w**
1. Fail to hold short as required (Taxi-out). Digital Copilot provides an auditory alert if the pilot is approaching a hold short line when not cleared to cross and it appears that the aircraft is not going to stop.	0.23
2. Fail to follow proper taxi route to proper runway (Taxi-out). Digital Copilot captures pilot readback of taxi instructions, and provides turn by-turn taxi instructions using auditory commands and visual guidance, including an auditory alert if the pilot deviates from the taxi route.	0.09
8. Fail to manage systems and maintain control including wake turbulence avoidance (Takeoff). Digital Copilot provides the pilot with a takeoff checklist and reads the checklist items aloud.	0.42
10. Fail to reject takeoff if insufficient runway remaining (Takeoff). Digital Copilot audibly reports the runway length remaining during the takeoff roll when there are 2000 feet or fewer remaining (but does not advise the pilot on whether to abort takeoff).	0.27
11. Fail to maintain proper climb profile or heading (Climb). Digital Copilot provides audible alerts if the climb rate (vertical speed) slows or if the aircraft begins to sink.	0.42
14. Fail to conform to flight plan / airspace requirements (En route).	

Digital Copilot provides an audible alert if the flight route deviates from way points entered by the pilot. In controlled airspaces, Digital Copilot captures pilot readback of altitude clearances and audibly alerts the pilot of clearance deviations. If the pilot sets an altitude, Digital Copilot will audibly alert the pilot of a deviation, and if the aircraft is approaching rising terrain, Digital Copilot alerts the pilot when vertical speed is insufficient to clear the terrain.	0.27
16. Fail to avoid hazardous weather (En route). Digital Copilot assimilates data from a variety of weather reports and monitors weather along the flight route, alerting the pilot to unfavorable trends over time (but does not help plan routing around weather). If the pilot modifies the route to avoid weather, and enters the revised route into Digital Copilot, the system will audibly alert the pilot of deviations from this revised routing.	0.28
17. Fail to avoid loss of control in hazardous weather (En route). Digital Copilot monitors for "slow rolls" where the bank angle steadily increases, and audibly alerts the pilot if this condition is detected.	0.28
18. Fail to manage systems and maintain control including fuel sufficiency (En route). Digital Copilot monitors for "slow rolls" where the bank angle steadily increases, and audibly alerts the pilot if this condition is detected. If requested by the pilot, Digital Copilot can also provide an audible reminder to switch fuel tanks (but does not monitor fuel levels or warn of potential fuel insufficiency).	0.45
19. Fail to avoid hazardous weather at destination (En route). Digital Copilot monitors weather conditions at the destination and audibly alerts the pilot if visibility or the ceiling fall below certain thresholds. Digital Copilot provides an alert when the weather changes from VMC to Marginal, from Marginal to IMC, or from IMC to low IMC (but does not assist in selecting a divert location or revising the flight plan).	0.26
20. Fail to avoid loss of control in hazardous weather at destination (En route). Digital Copilot monitors for "slow rolls" where the bank angle steadily increases, and audibly alerts the pilot if this condition is detected.	0.42
21. Fail to maintain proper descent profile or heading (Descent). Digital Copilot alerts the pilot if the aircraft is trending toward the ground (other than the airport runway surface).	0.48
24. Fail to conform to ATC directives at controlled airport (Approach). Digital Copilot automatically provides the tower frequency in an audible notification, and alerts the pilot if the tower will be open at the estimated arrival time. Digital Copilot provides guidance on joining and conforming to the approach pattern, by drawing the location of where to join the pattern, providing guidance on reaching that location, and alerting the pilot when the location has been reached. Digital Copilot also monitors conformance to the downwind leg and alerts the pilot if the aircraft is drifting toward or away from the runway. While monitoring conformance along the downwind and base legs, Digital Copilot alerts the pilot if an overly steep turn to final would be required or if predicted to pass through the final approach path. On approach, Digital Copilot alerts the pilot if the aircraft will pass through the extended runway centerline.	0.36
25. Fail to ID proper runway / approach at non-controlled airport (Approach). Digital Copilot provides guidance on the proper runway to use based on winds, and alerts the pilot if approaching a runway that is not open or if approaching the incorrect destination airport. Digital Copilot also provides other guidance and alerts in the approach phase of flight as described for action #24 above.	0.37
27. Fail to detect infeasible approach and execute go around (Approach). Digital Copilot continually monitors conformance to the approach pattern, providing guidance and alerts as described for actions #24 and #25 above (but does not advise the pilot on whether to go around or in executing a go around).	0.39
28. Fail to manage systems and maintain control including wake turbulence avoidance (Approach). Digital Copilot provides a GUMPS (Gas, Undercarriage, Mixture, Propeller, Seat belts and Switches) notification that alerts the pilot to check these items before landing. The system can also be configured to provide additional before-landing checklist items.	0.54
34. Fail to go around if insufficient runway remaining (Landing). Digital Copilot audibly reports the runway length remaining during landing (but does not advise the pilot on whether to go around).	0.35
35. Fail to hold short as required (Taxi-in). Digital Copilot provides an auditory alert if the pilot is approaching a hold short line when not cleared to cross and it appears that the aircraft is not going to stop.	0.23
36. Fail to follow proper taxi route from proper runway (Taxi-in). Digital Copilot captures pilot readback of taxi instructions, and provides turn by-turn taxi instructions using auditory commands and visual guidance, including an auditory alert if the pilot deviates from the taxi route.	0.09

III. Results

As described in Section II, HEPs needed for input to our PRA models are computed in two ways. First, in a baseline model without pilots' use of Digital Copilot, the HEP for each pilot action is quantified as the product of a nominal HEP (K, R, or S) and PSF multipliers s*c*t corresponding to PSFs of Stress/Stressors, Complexity and Available Time. Then, in a revised model, each HEP is multiplied by an additional factor w that accounts for the impacts of Digital Copilot on pilot Workload quantities and error probabilities. Results from these two models are presented in two subsections below, followed by a third subsection presenting results for the risk importance of each individual pilot action and associated features of the Digital Copilot.

A. Baseline Model Without Digital Copilot

In the baseline model, numerical PSF multipliers for Stress/Stressors (s), Complexity (c), and Available Time (t) were applied to HEPs (K, R, and S) appearing in PRA accident sequence equations. Numerical values for nominal HEPs were obtained by treating K, R, and S as variables in these accident sequence equations, which were aligned to historical accident frequencies in 14 discrete bins, summarized in Table 5. Historical data for each bin were obtained as a five-year average of accident rates derived from Nall reports [1-5] of general aviation accidents. These data, which included all pilot-related accidents for domestic, non-commercial, fixed-wing aircraft operating under Visual Flight Rules, accounted for over 70% of accidents that occur annually in general aviation. Using the historical data, along with the PSF multiplier values for s, c, and t in Table 3 and world event probabilities in Table 2, the 14 accident sequence equations were each solved for the value of one variable (K, R, or S). The resulting eight values for K were then averaged to obtain a point estimate for K, and similarly for the three values of R and three values of S, thereby obtaining the following point estimates:

K = 3.0E-04
R = 1.5E-06
S = 2.5E-07.

Table 5 Listing of 14 accident sequence bins

Event Tree	Accident Bin	Action Type
Taxi-out/in	a. Collision with Hazard	K
Takeoff	b. Collision with Hazard	K
Takeoff	c. Insufficient Runway	K
En route	d. Weather Related	K
Approach	e. Collision with Hazard	K
Approach	f. Failed Approach / Go-around	K
Landing	g. Collision with Hazard	K
Landing	h. Insufficient Runway	K
Takeoff	i. Loss of Control	R
En Route	j. Fuel Management	R
Approach	k. Loss of Control	R
Taxi	l. Loss of Control	S
Climb	m. Loss of Control	S
Landing	n. Loss of Control	S

To gauge the fit of the overall model to historical data, these point-estimate values for K, R, and S were then used in the 14 accident sequence equations to compute the annual number of accidents in each bin. The fit to the five-year historical accident rate data, illustrated in Fig. 2, is within about 20% for each bin and about 5% over all bins. For most bins, the fit is within the error bar that represents plus or minus one standard deviation in the five-year data. Over all bins, the total number of accidents annually is 767 for the model versus 734 for the data.

This empirical benchmarking of the baseline model is important because most inputs to the model, including PSF multipliers in Table 3 and world event probabilities in Table 2, were based on SME judgments. The good fit of model to data, illustrated in Fig. 2, supports the validity of these SME judgments as incorporated into the integrated PRA model of general aviation risk (without Digital Copilot).

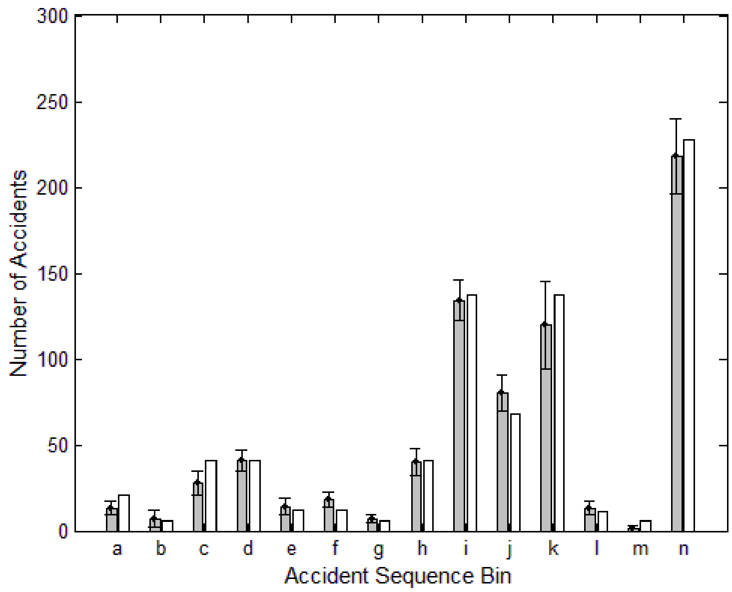

Fig. 2 Number of accidents annually in each accident sequence bin, model (white) and data (gray)

B. Revised Model With Digital Copilot

After benchmarking the baseline model (without Digital Copilot) against historical data, as reported above, a revised model was used to account for the expected impacts of Digital Copilot on pilot workload, error probabilities, and accident frequencies. This was done using pilot judgements of expected Workload without Digital Copilot (W) and with Digital Copilot (W'), to compute $\Delta W = W - W'$ along with a corresponding HEP multiplier (w) for each action via the Workload-Reliability Correlation.

As described in Section II, ratings of expected Workload were obtained from a total of 26 general aviation pilots, after the pilots reviewed a training briefing that provided detailed descriptions of Digital Copilot functionality. Each pilot provided ratings first assuming flight using a familiar Electronic Flight Bag without Digital Copilot technology, and then assuming flight using Digital Copilot technology integrated into the same Electronic Flight Bag. On average across all 26 pilots and the 19 actions affected by Digital Copilot, $\Delta W = 2.2$ on the scale of 0-20. The degree of inter-rater agreement was computed using a standard statistic known as the Intraclass Correlation Coefficient (ICC) [19]. Across all 26 pilots and the 19 actions affected by Digital Copilot, ICC(ΔW) = 0.85. In accordance with standard guidelines, ICC > 0.75 is considered excellent.

For each of the 19 actions affected by Digital Copilot, the distribution of ratings across all 26 pilots was used to obtain four different point estimates of ΔW, namely the mean (average), median (50th percentile), lower quartile (25th percentile), and upper quartile (75th percentile). Each set of point estimates, for all 19 actions, was then used in the WORC equation $w = e^{-0.56*\Delta W}$ to compute multipliers that account for the impacts of Digital Copilot on baseline HEPs. Finally, using these HEP multipliers, the PRA model was re-quantified and compared to results of the baseline PRA model (without Digital Copilot) that had been benchmarked earlier against historical accident rate data.

The resulting values of w, computed from mean values of ΔW across pilots, appear in Table 4. Figs. 3 and 4 compare the results of the PRA model using these HEP multipliers to the baseline model without pilots' use of Digital Copilot. Fig. 3 presents results for total accidents in all 14 bins. Fig. 4 presents results for fatal accidents, using a fraction for fatal/total accidents in each bin based on the five-year average of historical data [1-5].

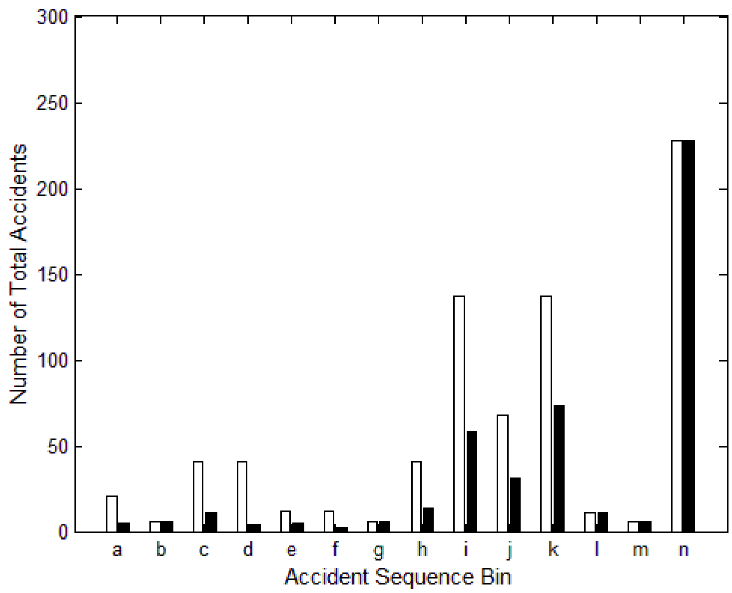

Fig. 3 Total accidents in each accident sequence bin, with Digital Copilot (black) and without (white)

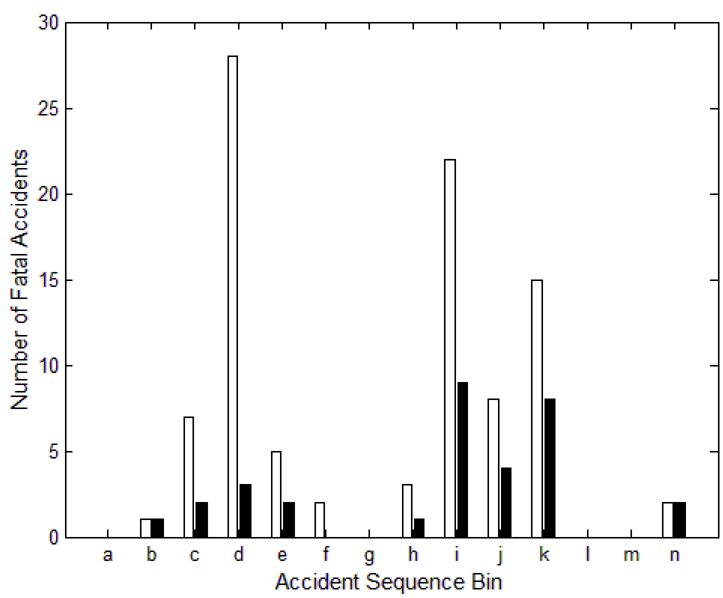

Fig. 4 Fatal accidents in each accident sequence bin, with Digital Copilot (black) and without (white)

Referring to bar heights in Fig. 3 with bin identifiers in Table 5, Digital Copilot has the largest impacts on total accident rates in bins i, j, and k involving Rule-based actions to maintain control during Takeoff, En route, and Approach phases of flight. Fig. 4 shows large impacts for fatal accidents in the same three bins, as well as a large impact in bin d involving Knowledge-based actions to avoid hazardous weather. Further insights into individual accident sequences within the 14 bins are provided by risk importance results presented below in Section III.C.

Over all accident sequence bins, assuming Digital Copilot technology is adopted and used by all pilots in general aviation, the model predicts an expected reduction of 40% in the total accident rate and 66% in the fatal accident rate. The significance of this reduction was established with a t-test, by comparing the predicted number of total accidents annually to the mean of the five-year historical distribution. This test showed that the expected number of total accidents with pilots' use of Digital Copilot was significantly less than the five-year mean, $p < 0.001$.

Besides these calculations using mean values of ΔW, the PRA model was also re-quantified using median, lower quartile, and upper quartile values of ΔW. The results of each case were compared to the baseline model (without pilots' use of Digital Copilot), to quantify the sensitivity of model results to variability in Workload ratings across pilots. The resulting percentages of total accidents and fatal accidents expected to be averted are summarized in Table 6, once again assuming Digital Copilot technology is adopted and used by all pilots in general aviation.

Table 6 Percentage of annual accidents expected to be averted by pilots' use of Digital Copilot

ΔW Values	**Total Accidents**	**Fatal Accidents**
Mean	40%	66%
Median	39%	64%
Lower Quartile	10%	25%
Upper Quartile	49%	77%

As seen in Table 6, the results are almost identical using mean and median values of ΔW, and these results for the mean and median differ more from the lower quartile than they do from the upper quartile. The reason is that ΔW values have an exponential impact on pilot error probabilities via the WORC equation $w = e^{-0.56*\Delta W}$. This exponential impact on error probabilities, in turn, produces a non-linear impact on accident sequence frequencies such that results tend to saturate as ΔW values increase from lower quartile to mean (or median) to upper quartile.

C. Risk Importance of Individual Pilot Actions

Besides results for 14 accident sequence bins and the overall sum of these bins, the PRA models developed in this study enable more detailed analyses of individual risk contributors. In particular, the notion of "risk importance" provides a direct comparison of risk contributed by each of the 38 pilot actions that appear as fault tree basic events in event tree accident sequences. Here risk importance is measured by Risk Reduction Worth (RRW), which is defined [20] as the reduction in accident frequency that would result if the human error probability for a pilot action was changed to zero (i.e., the potential for error was eliminated). Therefore, RRW represents the contribution of an individual error rate to the overall accident rate.

RRW was computed first for each action in the baseline model, without pilots' use of Digital Copilot, and then for each action in the revised model, with pilots' use of Digital Copilot. In each case, RRW was computed for total accidents and for fatal accidents as illustrated in Figs. 5 and 6, respectively. In these figures, the total (black + white) bar height for each action represents the risk importance (RRW) of that action in the baseline model without Digital Copilot, and the black bar height represents RRW in the revised model with Digital Copilot. Therefore, the white portions of bars represent how much risk has been reduced by Digital Copilot, and the black portions of bars represent how much risk remains with the Digital Copilot.

Referring to action #32 in Fig. 5, which has the highest RRW for total accidents, we see that RRW is the same both with and without Digital Copilot. The reason is that this is a Skill-based action to maintain control during landing, and Digital Copilot is an information system rather than a robotic control system that can assist the sensory-motor functions of landing an aircraft. However, loss of control during landing is rarely fatal, and Fig. 6 shows that action #32 has a small RRW for fatal accidents.

For other risk contributors to total accidents, Fig. 5 shows that Rule-based actions to maintain control during Takeoff (#8), En route (#18), and Approach (#28) phases of flight have relatively high RRW both with and without Digital Copilot. The white portions of bars show that the system has substantially reduced risk from these actions, and the black portions of bars show that substantial risk remains. For the same three Rule-based actions, Fig. 6 shows similar results for fatal accidents. Therefore, although the current system features are expected to accomplish risk reductions for these actions, the residual RRWs (black bars) suggest that the same actions remain candidates for additional system features to further reduce risk.

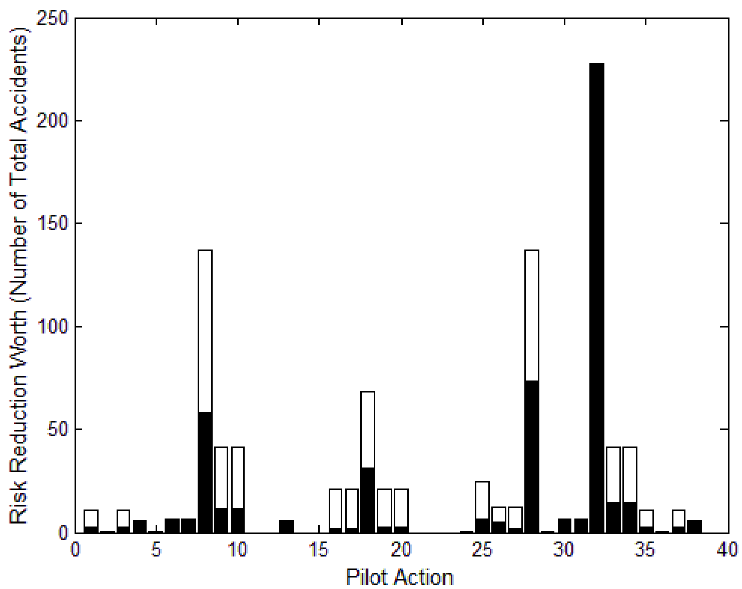

Fig. 5 Total Risk Reduction Worth for 38 pilot actions, with Digital Copilot (black) and without (white)

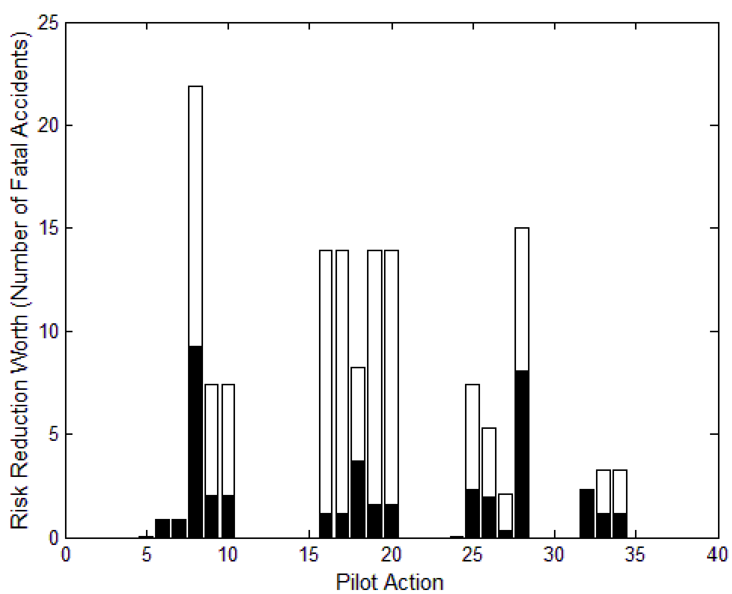

Fig. 6 Fatal Risk Reduction Worth for 38 pilot actions, with Digital Copilot (black) and without (white)

Besides Rule-based actions #8, #18, and #28, Fig. 6 shows that Digital Copilot is expected to achieve a large reduction in fatal RRW for several Knowledge-based actions including #16, #17, #19, #20 associated with avoiding hazardous weather. The residual RRWs (black bars) for these actions are small, which suggests that there is relatively little more risk to be averted by system features that might further address actions #16, #17, #19, and #20.

Regarding total accidents in Fig. 5, important Knowledge-based actions include #9, #10, #33, and #34, which are actions associated with ensuring sufficient runway during Takeoff or Landing. Digital Copilot directly affects Workload for actions #10 and #34, as noted in Table 4, but does not directly affect Workload for actions #9 or #33. However, Digital Copilot affects RRW for all of these actions because accidents result from failures of both #9 and #10 or both #33 and #34. That is, a reduction in error rate by Digital Copilot for action #10 reduces the RRW for action #9 appearing in the same accident sequence equation as #10, and a reduction in error rate by Digital Copilot for action #34 reduces the RRW for action #33 appearing in the same accident sequence equation as #34.

Taken together, these risk importance results are informing the continued development of Digital Copilot technology in efforts to reduce risk. Here it should be noted that these results are enabled not only by measures of pilot Workload and associated reductions in pilot error probabilities, but also by the integrated PRA models used to compute accident sequence frequencies. For example, based only on measures of Workload, the actions for which Digital Copilot achieves the greatest reductions are those with the lowest values of w in Table 4, namely Taxi actions #2 and #36. But Figs. 5 and 6 show that these actions have very small RRW even without Digital Copilot, so there is little risk to be averted regardless of how much the Workload may be reduced. In this case the benefits of the system are primarily associated with operational efficiency rather than operational safety. Conversely, Table 4 shows that action #28 to maintain control in Approach has the smallest Workload reduction (largest value of w) by Digital Copilot, yet Figs. 5 and 6 show that the associated risk reduction is among the highest of all actions modeled. Other actions for which Digital Copilot accomplishes relatively modest Workload reductions but large risk reductions include actions #8 and #18 to maintain control in Takeoff and En route, respectively.

IV. Conclusion

A. Contributions

This paper presented methods and results of analyses to quantify the safety benefits of a Digital Copilot in general aviation. The methods include PRA, HRA, and CPA. Our CPA is novel in its use of a Workload-Reliability Correlation, which overcomes limitations of the standard PSF approach to HRA in analyzing the effects of Ergonomics/Human-Machine Interface on HEPs.

An innovative aspect of our approach involved benchmarking the PRA model of accident sequence frequencies, including the HRA model of HEPs, against historical accident data. This enabled calculation of nominal HEPs for Knowledge-based, Rule-based, and Skill-based actions specific to the domain of general aviation. It also enabled benchmarking of PSFs for Stress/Stressors, Complexity, and Available Time, which are applied as multipliers to the nominal HEPs for individual pilot actions in specific situations. These benchmarked PSFs were then used to calibrate the new Workload-Reliability Correlation, which computes another multiplier to HEPs that accounts for the effects of Digital Copilot on pilot error probabilities, and thereby enables re-quantification of the PRA model to predict the expected impacts on accident sequence frequencies.

Taken together, our CPA, HRA, and PRA models predict an expected reduction in accidents assuming Digital Copilot technology is adopted by pilots in general aviation. Results were also computed for the risk importance of individual pilot actions, to identify those actions that contribute most to total and fatal accident rates, both with and without Digital Coplot. The detailed results for these pilot actions and associated system features are currently being used as a risk-informed basis for implementing Digital Copilot technologies and prioritizing ongoing development activities.

B. Limitations

The most significant limitation of our approach lies in the Workload ratings needed as input to the Workload-Reliability Correlation. These ratings are an improvement over the standard method of HRA, which uses judgments of a PSF for Ergonomics/Human-Machine Interface obtained from one or a few analysts (not operators) making only a categorical distinction between "Nominal" or "Good". Our Workload ratings were integer values in the range 0-20, obtained from a total of 26 general aviation pilots, but these ratings are still human judgments and therefore subject to biases and differences between raters. The ratings were found to exhibit excellent inter-rater reliability, but PRA results for accident sequence frequencies were sensitive to whether the model was quantified with mean (median), lower quartile, or upper quartile values from the distributions of ratings across pilots.

Regarding potential biases in ratings of Workload, it is important to acknowledge that the present study obtained pilots' judgments of *expected* Workload, not *experienced* Workload. In so doing, we were careful to explain the pilot actions of interest and describe associated features of Digital Copilot, to ensure that pilots had a credible basis for making judgments of Workload with and without the system. But these judgments were still based on *expectations* of pilots, rather than *experiences* of pilots using the prototype system in actual operations or flight simulations. Pilots may have been biased toward overestimating Workload reductions from Digital Copilot, due to wishful thinking about potential benefits or failure to imagine potential negative impacts of the system. Or, pilots may have been biased toward underestimating Workload reductions, because they had not actually experienced system benefits directly.

One direction for future research would be to obtain pilots' ratings of *experienced* Workload using Digital Copilot while engaged in flight simulations, and thereby benchmark the *expected* Workload ratings of the present study. But experienced Workload is still subjective and will vary between pilots and between flights for the same pilot, as well as between flight simulations and actual operations. Therefore, uncertainty in Workload ratings and variability in risk results would remain even if such further research were undertaken. Meanwhile, the present study based on pilots' ratings of expected Workload offers useful insights into the potential safety benefits of the prototype Digital Copilot. These insights include overall estimates of accident rates, as well as more detailed results for the relative risk importance of individual pilot actions and associated system features. The results for relative risk importance are arguably less susceptible to potential bias in pilots' ratings of expected Workload, because the same bias (optimistic or pessimistic) would presumably apply across pilot actions and system features.

Disclaimer

The contents of this document reflect the views of the authors and The MITRE Corporation and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

References

[1] 25th Joseph T. Nall Report, "General Aviation Accidents in 2013", Aircraft Owners and Pilots Association Air Safety Institute, Frederick, MD, 2016.

[2] 24th Joseph T. Nall Report, "General Aviation Accidents in 2012", Aircraft Owners and Pilots Association Air Safety Institute, Frederick, MD, 2015.

[3] 23th Joseph T. Nall Report, "General Aviation Accidents in 2011", Aircraft Owners and Pilots Association Air Safety Institute, Frederick, MD, 2014.

[4] 22th Joseph T. Nall Report, "General Aviation Accidents in 2010", Aircraft Owners and Pilots Association Air Safety Institute, Frederick, MD, 2012.

[5] 21th Joseph T. Nall Report, "General Aviation Accidents in 2009", Aircraft Owners and Pilots Association Air Safety Institute, Frederick, MD, 2011.

[6] Helmreich, R., Merritt, A., Wilhelm, J., "The Evolution of Crew Resource Management Training in Commercial Aviation", *The International Journal of Aviation Psychology*, Vol. 9, No. 1, 1999, pp. 19-32.

[7] Estes, S., Burns, K., Helleberg, J., Long, K., Pollack, M., and Stein, J., "Digital Copilot: Cognitive Assistance for Pilots", *Proceedings of the AAAI Fall Symposium on Cognitive Assistance in Government and Public Sector Applications*, Arlington, VA, Nov. 2016.

[8] Burns, K., Bonaceto, C., Estes, S., and Helleberg, J., "Evaluating the Operational Safety of a Digital Copilot", *Proceedings of the AAAI Fall Symposium on Cognitive Assistance in Government and Public Sector Applications*, Arlington, VA, Nov. 2017.

[9] Stamatelatos, M., and Homayoon, D., "Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, Second Edition", NASA/SP-2011-3421, National Aeronautics and Space Administration, Washington, DC, 2011.

[10] Vesely, W., Goldberg, F., Roberts, N., and Haasl, D., "Fault Tree Handbook", NUREG-0492, US Nuclear Regulatory Commission, Washington, DC, 1981.

[11] SAPHIRE, Systems Analysis Programs for Hands-on Integrated Reliability Evaluations, Version 8, NUREG/CR-7039, US Nuclear Regulatory Commission, Washington, DC, 2011.

[12] Gertman, D., Blackman, H., Marble, J., Byers, J., and Smith, C., "The SPAR-H Human Reliability Analysis Method", NUREG/CR-6883, US Nuclear Regulatory Commission, Washington, DC, 2005.

[13] Rasmussen, J., "Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, 1983, pp. 257-266.

[14] Hart, S., and Staveland, L., "Development of the NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research", *Human Mental Workload*, edited by P. Hancock and N. Meshkati, Elsevier, Amsterdam, 1988, pp. 139-183.

[15] Reid, G., and Nygren, T., "The Subjective Workload Assessment Technique (SWAT): A Scaling Procedure for Measuring Mental Workload", *Human Mental Workload*, edited by P. Hancock and N. Meshkati, Elsevier, Amsterdam, 1988, pp. 185-218.
[16] Vidulich, M., and Tsang, P., "Absolute Magnitude Estimation and Relative Judgment Approaches to Subjective Workload Assessment", *Proceedings of the Human Factors Society 31st Annual Meeting*, Human Factors Society, Santa Monica, CA, 1987, pp. 1057-1061.
[17] Estes, S., "The Workload Curve: Subjective Mental Workload", *Human Factors*, Vol. 57, No. 7, 2015, pp. 1174-1187.
[18] Rasch, G., *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen, Danish Institute for Educational Research, University of Chicago Press, Chicago, 1980.
[19] Shrout, P., and Fleiss, J., "Intraclass Correlations: Uses in Assessing Rater Reliability", *Psychological Bulletin*, Vol. 86, No. 2, 1979, pp. 420-428.
[20] Vesely, W., Davis, T., Denning, R., and Saltos, N., "Measures of Risk Importance and Their Applications", NUREG/CR-3385, US Nuclear Regulatory Commission, Washington, DC, 1983.