AIAA AVIATION Forum
17-21 June 2019, Dallas, Texas
AIAA Aviation 2019 Forum

10.2514/6.2019-3443

Assessing Potential Biases in Risk Perception for General Aviation Pilots

Nicoletta Fala[1] and Karen Marais[2]
Purdue University, West Lafayette, IN, 47907

In 2017, General Aviation (GA) in the United States was responsible for 990 non-commercial fixed-wing accidents—157 of them fatal. One way to improve safety is through improved training and flight visualization. The introduction of Electronic Flight Bag (EFB) applications to the flight deck has allowed post-flight visualization and debrief tools to emerge in an attempt to provide pilots with more information about their flights. While we could assume that safety-driven post-flight feedback that alerts pilots to potentially hazardous situations can facilitate risk management in subsequent flights, it is not clear how the way we present feedback affects how pilots perceive it, or what the best way to present it is. In this paper, we develop different ways of presenting pilots with safety-driven information, informed by the literature on risk assessment and cognitive biases, and design a survey to identify potential implications of communicating risk to pilots in different ways.

I. Nomenclature

CFI	=	Certified Flight Instructor
FAA	=	Federal Aviation Administration
FAR	=	Federal Aviation Regulation
FDR	=	Flight Data Recorder
FOQA	=	Flight Operations Quality Assurance
GA	=	General Aviation
NTSB	=	National Transportation Safety Board

II. Introduction and Motivation

Even though General Aviation (GA) safety has improved over the past years, several hundred pilots still lose their lives in GA accidents yearly. In 2017, GA in the United States was responsible for 1193 accidents—185 of them fatal [1]. Non-commercial fixed-wing GA has been contributing disproportionately to the aviation safety record; its accident rate of 5.73 accidents per 100,000 flight hours is significantly higher than the rate of 2.03 accidents per 100,000 flight hours in commercial fixed-wing GA [1]. With GA having such a large presence in the national airspace, safety is a pressing concern.

One potential way to improve GA safety is to continue providing pilots with feedback on their flying even after they finish their training and are no longer flying with an instructor (and potentially flying as the sole pilot or the most experienced pilot in the aircraft). Commercial products, such as CloudAhoy and CirrusReports, take advantage of the addition of technology in the cockpits of small aircraft to collect flight data. Such products then use the flight data they collected to create a post-flight debrief. This debrief provides pilots with feedback using a visualization of their flights but does not discuss risk or flight safety. To manage risk, however, pilots must first perceive the risk associated with a situation or hazard, and then decide whether they are willing to accept this amount of risk in this situation [2]. Safety-driven post-flight feedback may help facilitate risk management in subsequent flights by alerting pilots to potentially hazardous situations. Feedback must help pilots perceive risk correctly and encourage them to reduce risk by improving their flying and changing any potentially unsafe behaviors.

In past research, we used flight data from a Garmin G1000 to detect unsafe behaviors [3], or *hazardous states* [4]. We used the NTSB database of accidents to generate a list of hazardous states or events that could potentially put the safe outcome of a flight at risk and mapped them to parameters that we calculate using available data sources (see Table 8 in the Appendix). Once we detect a *hazardous state* in a flight [3], we need to inform the pilot so that they can change their behavior. We hypothesize that we can impact how pilots respond to feedback through the way we

[1] Graduate Research Assistant, School of Aeronautics and Astronautics, 701 W. Stadium Ave, Student Member.
[2] Associate Professor and Associate Head for Undergraduate Education, School of Aeronautics and Astronautics, 701 W. Stadium Ave, Associate Fellow.

show them their flight data. To test this hypothesis, we designed and disseminated a survey that aimed to investigate which cognitive biases pilots are susceptible to and how to best communicate data-driven risk feedback to GA pilots.

In this research, we aim to help pilots improve the safety of their flights by providing them with data-driven risk information in the form of post-flight debrief. Designing risk messages to either take advantage of potential biases or circumvent them could help flight training professionals motivate behavioral changes among students and new pilots. Section III of this paper reviews the literature on cognitive biases that affect pilot risk perception, and Section IV describes a survey we designed to test for the impact of these cognitive biases in a sample of approximately 270 pilots. Section V discusses and analyzes the results of the survey, breaking it down in three factors: framing language, parameter type, and representation method. Section VI concludes the paper and discusses the future work in this research.

III. Cognitive Biases in Risk Communication

Cognitive biases affect the way people communicate, process information, and make decisions in all industries. Understanding these biases is important in using communication to achieve a mission. In flight training, the mission is usually to help pilots fly more safely or more efficiently. For example, flight instructors and examiners expect pilots to take off at the appropriate airspeed and fly to their destination in a straight line with as few inadvertent deviations as possible. Intentional or inadvertent risky behaviors by pilots is one of the major causes of GA accidents, with pilot error appearing at the top of accident factors annually [1]. In many cases, pilots may not feel that what they are doing is as risky as their flight instructor tells them, or, they may not be aware that what they are doing is risky at all. For example, an excessive bank angle in the pattern while at a low airspeed can induce an unrecoverable stall, especially on the base to final turn. However, inexperienced pilots will still try to avoid having to go around by steepening their last turn to make their landing. Better and continuous feedback about risk may help reduce these behaviors. Flight instructors and commercial debrief products can increase the effectiveness of communication when trying to mitigate such accidents by taking advantage of cognitive biases. Past research on cognitive biases has focused on the general population as a whole (for example, in conveying health risks to patients [5, 6, 7]) or specific sub-populations (for example, the student athlete population in sports coaching), but not aviation [8].

The mission of debrief messages in aviation is similar to the missions of other areas studied—physicians are trying to convince patients to alter specific behaviors that could be potentially hazardous to their health, sports coaches are trying to motivate their athletes to improve their performance, and flight instructors or managers are trying to convince their students or employees to change behaviors that are unsafe or inefficient. Prior literature is therefore useful in generating a list of biases that could influence pilots. However, pilots are not necessarily a representative sample of the population—studies on pilot personalities have identified that pilots tend to be highly conscientious, deliberative, achievement striving, highly competent, and dutiful [9]. The strong presence of specific traits among pilots suggests that they may also be affected differently than the general population. All pilots are also required to have a degree of numeracy and literacy, since they need to prove their knowledge and application of numerical concepts in a written test—an assumption that likely does not apply to the general population.

In our review of the literature, we focused on the cognitive biases that affect how people perceive risk. Specifically, we consider framing techniques in language and measurements, and the actual representation method used. In this paper, we evaluate three potential biases based on our review of the literature: the framing language, the type of parameter, and the representation method. Language can be safety-centric or risk-centric—a flight that was very safe is also not risky at all. Parameters that characterize behaviors can be categorized in terms of safety or performance. Consider for example *deviation from the centerline*—flight instructors expect their students to land on the centerline during every landing and stay on the centerline while taking off. On a very narrow runway, maintaining the centerline at all times is important in avoiding runway excursions. In contrast, on a very wide runway, it is hard to make the case that deviating from the centerline by a small distance is unsafe, if the aircraft was nowhere near the edge of the runway. *Distance from the centerline* is a performance parameter—minimizing it takes the flight closer to perfection. *Distance from the runway edge* is a safety parameter—minimizing it minimizes safety. Lastly, the unsafe behavior can be represented graphically or numerically.

IV. Survey Design Considerations

To determine whether pilots are susceptible to some of the biases described in Section III, we designed and disseminated a survey that asked pilots to debrief hypothetical flights. Our survey uses modified CloudAhoy screenshots from the takeoff phase of three flights to help pilots visualize the flight, where we added information on risk, before asking them to assess the risk. We want safety-driven debrief to 1) communicate to pilots how risky their flight was, and 2) if they displayed unsafe behaviors in their flying, motivate a change to happen.

A full factorial design is appropriate for evaluating the three potential biases and their interactions. Our three potential biases correspond to three factors, and they each can take two levels, making this a 2^3 full factorial experiment. Different combinations of the levels of each factor result in eight treatments, as shown in Table 1. The full factorial design also allows us to test for potential interactions between factors. For example, changing the language used to frame the risk and the representation method independently might not affect how pilots perceive their risk, but a particular combination of the two factors might change risk perception. To test all three factors and their interaction effects, we disseminated eight versions of our survey—one for each treatment[3].

Table 1: The eight treatments represent the different ways we used to show pilots their debrief data. Under language, "+" represents risk-centric language and "-" represents safety-centric language. Within parameter type, "+" represents performance-driven parameters, and "-" represents safety-driven parameters. Graphical representation is represented by "+" and numerical representation by "–".

Treatment	Language	Parameter	Representation
1	+	+	+
2	+	+	–
3	+	–	+
4	+	–	–
5	–	+	+
6	–	+	–
7	–	–	+
8	–	–	–

To evaluate both pilots' risk perception and their motivation to change behaviors that they deemed unsafe, we ask:

1) Given the information presented to you, how safe would you say this takeoff was? (Scale of 1 to 5, with 1 being very unsafe and 5 being very safe.)
2) In this takeoff, which of the following would concern you? (Centerline deviation, Rotation airspeed, Engine RPM, Takeoff distance, Wind)
3) What changes do you think you could make to an upcoming flight after the information presented here, if any?
4) How likely are you to make each of these changes to an upcoming flight?
5) How important do you think each of these changes is to improving safety on takeoff?

The first two questions measure the respondents' risk perception. Changes in the average response to the first question may indicate that one representation method is pushing pilots towards thinking the flight is either safer or riskier. The second question is more objective in nature—we compare the survey responses to the flight data to identify whether some representation methods resulted in the pilots ignoring or missing a behavior that was actually present in the data. The last three questions aim to measure willingness to change. We assess willingness to change based on factors like the number of changes suggested, the maximum likelihood of making a change, out of the changes suggested, and the likelihood of making the change that scored the highest on the last question (i.e., the change the respondent believes is the most important).

In previous work we identified the events or behaviors that appear most frequently in NTSB reports of takeoff accidents [3, 10]. Airspeed and engine power at takeoff affect aircraft performance during the rotation and initial climb phases of flight. Insufficient airspeed at rotation may result in the airplane settling back down on the runway or stalling with inadequate altitude for the pilot to respond appropriately. The wind conditions for the runway of use also affect performance, and depending on the pilot's experience level, can be dangerous. The amount of crosswind one can take off in is dependent on the airplane, but also on how much experience the pilot has in similar crosswind conditions. Tailwind takeoffs can be dangerous even with small tailwinds. The tailwind increases groundspeed, which in turn

[3] The full survey is available at www.nicolettafala.com/survey.

increases takeoff distance and delays rotation, leaving the pilot with potentially insufficient time or distance to climb at a vertical speed that will ensure obstacle clearance. We are also concerned about deviation from the runway's centerline and how far down the runway the airplane ends up taking off. These behaviors should appear in flights infrequently, if pilots are flying safely. We developed algorithms to calculate and measure all of these events using flight data. Table 8 in the Appendix explains how we calculate the required parameters for one of the events, *Deviation from centerline*. We then supplemented the CloudAhoy screens from the same flight data with risk information based on the output of the algorithms, presented in the different ways in Table 1 [11].

To avoid introducing different biases than the ones we are testing for in the survey responses, we need the respondents to have an equal probability of being randomly assigned a treatment out of the eight possible treatments. Since we are testing for eight treatments, we need a large number of responses—for a power of 0.90, we need 22 responses for each treatment to be able to detect a maximum difference of 0.5 between the mean response to each treatment. A web-based survey has the potential to collect data from a large and dispersed sample of participants [12], providing us with access to individuals in distant locations or participants who may be difficult to reach [13]. However, self-selection bias may result in a systematic bias—some individuals are more likely than others to complete the survey, while others will tend to ignore the invitation to participate in the online survey. Such sampling issues potentially inhibit our ability to generalize and estimate population parameters, meaning that we can only use our results as an indication of a potential bias in the population that is only definitely present in the sample we surveyed.

The survey asked the pilots to debrief three flights (Flight A, Flight B, and Flight C). The data for the three flights was recorded by a Garmin G1000 display on a Cessna 172. All three flights took off from the same runway at OSU (The Ohio State University) and flew a rectangular pattern over Columbus before returning to the airport. However, the three flights differ in their takeoff risk: Flight A had the riskiest takeoff out of the three, and Flight B the safest. Every respondent went through a tutorial to become comfortable with using the interactive debrief tool (static example shown in Figure 1), and then debriefed one flight at a time and answered questions. In this paper, we will use the responses to the questions to measure feedback effectiveness in terms of risk perception and motivation to make behavioral changes.

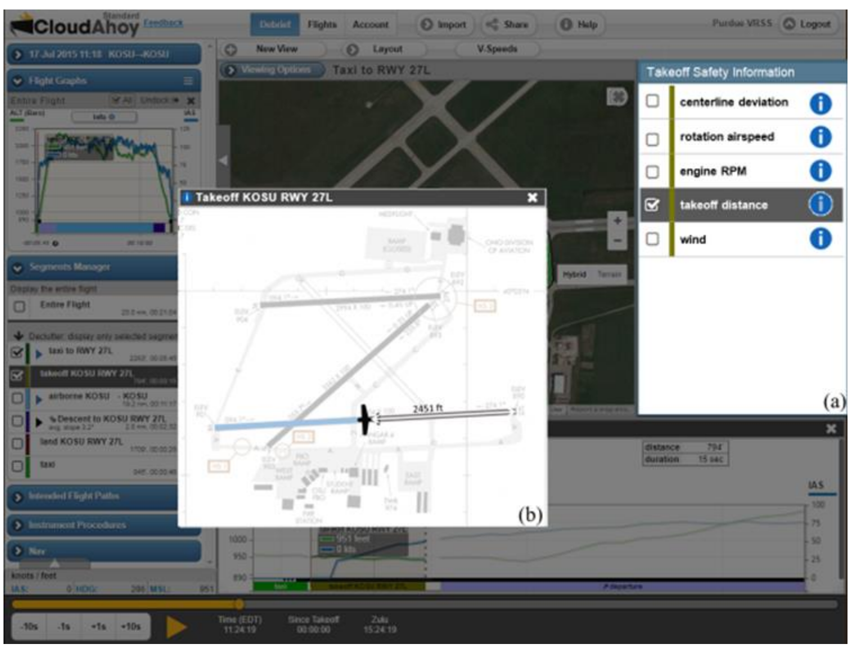


Figure 1: We supplemented the visualization of the flight data from CloudAhoy with information on the safety of the flight on five parameters to create an interactive prototype[4] that survey respondents could use to debrief sample flights [11].

[4] An example of the interactive debrief prototype is available at www.nicolettafala.com/debriefexample.

To disseminate the survey, we used various aviation groups and newsletters, such as the Curt Lewis and Associates Flight Safety Information newsletter (distributed daily to more than 36,000 subscribers and tailored to people with an interest in aviation safety), flight clubs, and flight schools. We also used social media to distribute the survey to those pilots who may not be a part of type clubs or other organizations and encouraged snowball sampling. Out of the approximately 200 respondents who completed the demographics part of the survey, 71% were male and 25% female, and most of them had at least a college degree. Most respondents held a private license (49%) or a commercial license (30%) and we observed a high variability in their occupation. The least experienced pilot in our sample had 18 flight hours, and the most experienced pilot had 28,000 flight hours, with most respondents flying at least once a week (54%). Surprisingly, 89% of all respondents had not used aviation debrief products before, so their interaction with our survey was their first introduction to products similar to CloudAhoy.

V. Results and Discussion

In this section, we discuss the analysis of the survey data for the three factors individually (framing language, parameter type, and representation method). We present the descriptive statistics for each factor and test for differences in the distributions of the risk perception and motivation to change behavior distributions using the Mann-Whitney-U test.

A. Framing Language

Framing language refers to how pilots responded when asked "*How risky would you say this takeoff was?"* versus "*How safe would you say this takeoff was?*" While mathematically a flight that is not too risky is by definition very safe, the phrasing did affect how pilots perceived the risk. Although Flight C was not affected by the framing language, in the safety-centric language there is a higher concentration towards the neutral value. However, the framing language factor changed the mode in Flight B and altered the distribution in Flight A. Flight B also shows a general movement towards the riskier side when using a safety-centric framing language. The impact on Flight A was different than the impact on Flight B—in the first case, framing the question in safety-centric language resulted in more people saying the flight was not safe, whereas in the second case framing the question in risk-centric language decreased the number of people saying the flight was not risky and increased the number of people reporting that the flight was risky.

Table 2 shows some of the descriptive statistics on risk perception for the risk-centric and safety-centric framing languages among the three flights. The framing language resulted in a change in the mean risk perception in Flight A but did not affect Flights B or C.

Table 2: The framing language reduced the risk of the risk perception in Flight A.

	Risk rating							
	Safety-centric				Risk-centric			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	3.2519	1.0631	3	1.75	2.9398	0.9436	3	2
B	2.9802	1.0953	3	2	3.0106	1.2742	3	2
C	2.9897	0.9947	3	2	2.8696	0.9858	3	2

The distributions of the number of changes pilots suggested after reviewing their debrief for all three flights look the same for all flights independent of the framing language used. Table 3 shows the descriptive statistics for all three flights. The difference between the two framing languages is much lower in this factor than the *representation method* and *parameter type* factors.

Table 3: The framing language did not change the average number of changes that the pilots said they would make to an upcoming flight.

	Number of changes							
	Safety-centric				Risk-centric			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	1.5556	1.3308	2	2	1.4286	1.3833	1	3
B	1.1486	0.9098	1	2	1.2447	0.9121	1	1
C	1.4536	1.1816	2	2	1.4239	1.0611	1	1

B. Parameter Type

The second factor, *parameter type*, also take one of two values: safety parameters or performance parameters. *Parameter type* refers to whether the parameter is presented in terms of risk or performance. For example, comparing the amount of runway that the pilot used in taking off to the takeoff distance specified in the aircraft handbook, tells the pilot how close they were to the nominal way of flying in comparison to the handbook. Reporting the amount of runway that remained after takeoff instead (i.e., the runway length that was not used) aims to communicate how much room for error the pilot had, based on how close the pilot and aircraft are to an unsafe situation or incident. In the first case, the pilot should want to minimize the number; in the latter case, a higher number is better.

The distribution of risk perception responses in Flight A had a slightly smaller variance in the safety representation. The safety parameters in Flight B moved the responses to the right, towards "extremely risky," with a different mode in the safety and performance parameters cases. The *parameter type* made no noticeable difference in the means in Flight C. Both Flight A and C maintained the same mode. Table 4 shows the descriptive statistics on risk perception for the performance and safety parameter types among the three flights. The average risk ranking of the three flights also changed between *performance parameter* and *safety parameter*.

Table 4: Pilots risk perception changed between performance and safety parameter types in Flight B, with the mean increasing from 2.63 in the performance parameter type to 3.38 in the safety parameter type.

	Risk rating							
	Performance				Safety			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	3.1538	1.0299	3	2	3.0320	0.9995	3	2
B	2.6337	1.0145	3	1	3.3830	1.1742	3	1
C	2.9570	0.9659	3	2	2.9063	1.0165	3	2

The safety parameter increased the number of changes pilots suggested after reviewing their feedback in Flights B and C, but decreased the mean of the number of changes in Flight A, as shown in Table 5. For the riskier takeoff in Flight A, pilots suggested they would make more changes to an upcoming flight when presented with the performance parameters. The safety parameter version of the debrief reduced the number of respondents who opted to continue without making any changes in Flight B and C (the safer takeoffs) but increased the "no changes" responses in Flight A (the riskier takeoff).

Table 5: The numerical representation method resulted in a higher number of changes overall.

	Number of changes							
	Performance				Safety			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	1.6364	1.3612	2	3	1.3280	1.3367	1	2
B	1.000	0.8602	1	2	1.4043	0.9196	1	1
C	1.3226	1.0951	1	2	1.5521	1.1413	1.5	1

C. Representation Method

The *representation method* factor takes one of two levels: graphical, or numerical. The representation method factor changed the mode of risk perception only in Flight B. Flight C appeared to be largely unaffected. Flight A maintained the same mode, but the graphical representation was more uniform in distribution around the mid-point than the numerical level. Flights A and B seem to have moved in opposite directions—the graphical representation moved the responses slightly towards the riskier side in Flight A compared to the numerical representation, but distinctively towards the less risky side in Flight B. Table 6 shows the descriptive statistics on risk perception for the graphical and numerical representation methods among the three flights.

Table 6: The way pilots rated their risk changed between graphical and numerical representation methods. The largest change happened in Flight B, with the mean increasing from 2.76 in the graphical method to 3.20 in the numerical method.

	Risk rating							
	Graphical				Numerical			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	3.1951	0.9889	3	2	3.0138	1.0340	3	2
B	2.7582	1.2679	3	2	3.2019	1.0647	3	2
C	2.9277	0.9342	3	2	2.9340	1.0353	3	2

The number of changes the pilots suggested after reviewing their feedback is an indication of how motivated they are to change unsafe behaviors. Overall, the safer of the three flights resulted in a lower number of suggested changes, and the riskier flight resulted in pilots suggesting more changes they would make to an upcoming flight. Although the numerical representation did decrease the number of respondents who opted for no changes after their debrief in Flights B and C, the same did not apply to Flight A, the riskier flight. In Flight B, in particular, the numerical representation resulted in most respondents saying they would make two changes, whereas the mode for the graphical representation method was zero, as shown in Table 7.

Table 7: The number of changes pilots suggested after reviewing their debrief ranged from zero to five. The numerical representation method resulted in a higher number of changes overall.

	Number of changes							
	Graphical				Numerical			
Flight	Mean	Standard Deviation	Median	IQR	Mean	Standard Deviation	Median	IQR
A	1.3984	1.3474	1	2	1.5724	1.3629	2	3
B	1	0.9661	1	2	1.3654	0.8251	1	1
C	1.3133	1.1575	1	2	1.5377	1.0882	2	1

D. Flight Comparison

The framing language factor only impacted the risk perception metric in Flight A. Representation method impacted both the risk perception metric and the number of changes proposed in Flight B, but not in the other two flights. Parameter type changed the distribution of the number of changes proposed in Flights A and B and the risk perception metric in Flight B. There was a significant interaction effect between representation method and parameter type that affected the risk perception metric in Flights A and B.

The results were not the same for all three flights evaluated, suggesting that the flight may be a factor that impacts the type of feedback that would be most effective.

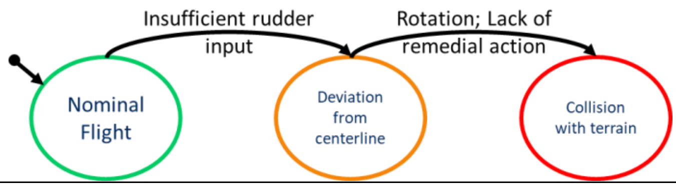
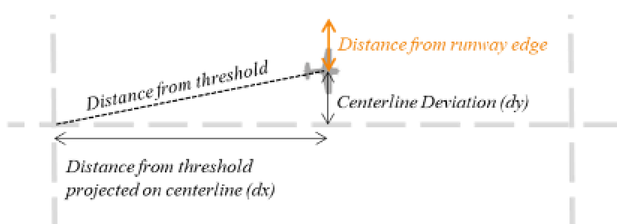
VI. Conclusion and Future Work

In this paper, we used a survey to evaluate the effectiveness of different risk communication methods developed in earlier work [11] in an attempt to identify the best way to communicate unsafe behaviors to pilots. We defined feedback effectiveness based on whether it communicated the risk of the situation and whether it motivated pilots to improve their flying by doing something to mitigate the unsafe behaviors. We created and disseminated a survey based on a full-factorial design to evaluate whether three factors we chose based on the literature on cognitive biases in risk communication in other fields affect feedback effectiveness in aviation in a similar way. We analyzed the results from 268 responses and showed that the feedback representation affects its effectiveness in terms of risk perception, but not the pilots' motivation to change. The effect of the three factors is not consistent across the three flights. The impact of the three factors on risk perception in different flights may be a function of the risk level of each flight or the kinds of hazardous states present in each flight. The next step in this work is to investigate the effect of the flight characteristics on risk perception.

In future work, we will expand this research to other phases of flight (approach, landing, cruise, etc.) by identifying sets of hazardous states that are important to communicate and developing algorithms to detect them in flight data. Additionally, we will investigate potential effects of demographics on the results. Specifically, we will evaluate the effect of gender, age, flight experience (certificate level and number of flight hours), flight training background (Part 61 versus Part 141), education level, and occupation.

Appendix

Table 8: Detecting the *Deviation from Centerline* Hazardous State in Flight Data

State Definition: Deviation from centerline	
Severity	Deviation from the runway is usually the result of insufficient rudder control while accelerating. As the pilot advances the throttle to full power, the left-turning tendencies of a single-engine airplane increase, requiring right rudder application to counteract them. Maintaining directional control on the runway is important both during the takeoff and landing phases.
Accident Example (**GAA16CA284**)	A pilot flying a Citabria, a tailwheel airplane, in Ferndale, MT, in 2016, drifted left of the runway centerline during his takeoff roll. He attempted to correct by applying right rudder, which resulted in the airplane slowing down, suggesting that the pilot was touching the brakes. The pilot released the right rudder to adjust his foot so that it would not touch the brake, and noticed that the airplane was quickly approaching the left edge of the runway. He decided to rotate early, but the airplane continued deviating towards the left, and ended up colliding with a hangar and catching fire. The NTSB reported that the cause of the accident was "the pilot's loss of directional control during takeoff, resulting in a decision to rotate early, and a collision with a hangar and subsequent fire."
Accident State-Based Model	
State Detection: Deviation from centerline	
Parameter	*Distance from centerline/Distance from runway edge*
Type	Multi-source parameter • Flight data (FDR, Smartphone, or ADSB) • National Transportation Atlas' Airport Runway Database (RITA, 2016)
Data required	From flight data: • GPS Coordinates at takeoff point From runway database: • GPS coordinates at the runway threshold on both ends of the runway • Runway width
Calculation	1. Identify the takeoff point in flight data 2. Detect the airport and runway from which the aircraft took off by finding airports/runways that fall in a boundary box around the takeoff point 3. Use GPS coordinates to calculate the distance between the takeoff point and the runway threshold 4. Use the coordinates of the threshold of the runway at the two ends to find the centerline 5. Project the distance from the threshold onto a line perpendicular to the centerline and correct for the Earth's curvature (*dy*) to obtain the *Distance from centerline* 6. Subtract the projected distance from the runway width to obtain the *Distance from runway edge* 
Risk levels for a Cessna 172	Risk level 1: *Distance from runway edge* > $0.75\left(\frac{Runway\ Width}{2} - 18\right)$ Risk level 2: *Distance from runway edge* > $0.5\left(\frac{Runway\ Width}{2} - 18\right)$ Risk level 3: *Distance from runway edge* > $0.25\left(\frac{Runway\ Width}{2} - 18\right)$

Acknowledgments

We thank the Purdue Statistical Consulting Service for their help with the data analysis, and the people and organizations that were significantly helpful in our snowball sampling (the Ninety-Nines and the Cardinal Flyers in particular).

References

[1] AOPA Air Safety Institute, "2016-2017 GA accident scorecard," 2018.

[2] Hunter, D. R., "Risk perception and risk tolerance in aircraft pilots," DOT/FAA/AM-02/17, 2002.

[3] Fala, N., and Marais, K., "Detecting Safety Events During Approach in General Aviation Operations," *16th AIAA Aviation Technology, Integration, and Operations Conference,* AIAA, Washington, DC, 2016..
doi: 10.2514/6.2016-3914

[4] Rao, A., "A New Approach to Modeling Aviation Accidents," Ph.D. Dissertation, School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN, 2016.

[5] Gurm, H. S., and Litaker, D. G., "Framing procedural risks to patients: Is 99% safe the same as a risk of 1 in 100?" *Academic Medicine.* Vol 75(8), 2000, pp. 840-842.

[6] Lipkus, I. M., "Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations," *Medical Decision Making.* Vol 27, 2007, pp. 697-714.
doi: 10.1177/0272989X07307271

[7] Paling, J., "Strategies to help patients understand risks," *BMJ*. Vol 327(7417), 2003, pp. 745-748.
doi: 10.1136/bmj.327.7417.745

[8] O'Hare, D., "Aeronautical Decision Making," In: P. S. Tsang & M. A. Vidulich, eds. *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Chap. 6.

[9] Fitzgibbons, A., Schutte, P., & Davis, D., "Pilot personality profile using NEO-PI-R," *National Aeronautics and Space Administration*, 2004.

[10]National Transportation Safety Board, "Aviation Accident Database & Synopses",
http://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx, Accessed November 2018.

[11]Fala, N., and Marais, K., "Communicating Data-Driven Risk Information to Pilots," *20th International Symposium on Aviation Psychology*, Dayton, OH, 2019.

[12]Leong, F. T. L., and Austin, J. T., *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants*, 2nd ed., SAGE Publications, Inc., 2006.

[13]Wright, K. B., "Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services," *Journal of Computer-Mediated Communication*, Vol 10(3), 2005.
doi: 10.1111/j.1083-6101.2005.tb00259.x