Retrospective and Prospective Validity of Aircraft Accident Risk Indicators

David R. Hunter, Federal Aviation Administration, Washington, D.C.

Data from a national survey of pilots were used to examine the validity of measures for the prediction of aviation accidents that had occurred prior to the survey (retrospective analysis) and accidents that occurred after the survey (prospective analysis). Separate retrospective and prospective analyses were conducted, and 45 measures from the survey were found to be associated significantly with accident involvement in the retrospective analysis. However, only 13 of those 45 measures achieved significance in the prospective analysis. Most of the measures found to be significantly related to accident involvement concerned aviation exposure; the remaining measures related to pilots' perceived and actual level of caution. The study is unique in its use of a cohort design for the examination of aircraft accident risk prospectively, and the results suggest the need for caution in the interpretation of retrospective analyses in this research domain. Actual or potential applications of this research include the design of aviation safety programs and the design or interpretation of studies that address indicators of aircraft accident risk.

INTRODUCTION

Numerous studies have examined the relationships among pilot biodata, personality measures, and accident involvement. Commonly examined characteristics include age (Baker, Lamb, Li, & Dodd, 1993; Lategola, Fiorica, Booze, & Folk, 1970; McFadden, 1997; Urban, 1984), gender (Harkey, 1997; Lubner, Markowitz, & Isherwood, 1991), education (Platenius and Wilde, 1989; Urban, 1984), attitudes toward flying (Platenius & Wilde, 1989), flying experience (Baker et al., 1993; Harkey, 1997; O'Hare, Chalmers, & Bagnall, 1996; Platenius & Wilde, 1989; Urban, 1984), and previous accident involvement (Baker et al., 1993). A recent comprehensive review of this literature is provided by O'Hare (1999).

In all these studies, retrospective study methods have been used, in which currently measured variables are used to predict a preceding accident occurrence. Prospective methods (sometimes referred to as *cohort studies* in the epidemiological literature) in which variables are used to predict a future accident occurrence have not been reported, although they would generally be regarded as a stronger form of validity.

Li (1994) reviewed epidemiologic studies of pilot-related factors in aircraft crashes. Commenting on study design, methods, and major findings, he noted, "The cohort study design is the only method allowing full investigation of the natural history of a phenomenon" (p. 947). However, Li found only three cohort studies conducted among pilots, none of which used pilot performance or accidents as the outcome of interest.

The present study uses data obtained from a national survey of pilots conducted in 1994 to examine the validity of measures for predicting accidents that had occurred prior to the survey (retrospective analysis) and accidents that occurred after the survey (prospective or cohort analysis). The survey included measures of most of the variables that had been found to

Address correspondence to David R. Hunter, Office of Aviation Medicine, Federal Aviation Administration, 800 Independence Ave., Washington, DC 20591. **HUMAN FACTORS,** Vol. 43, No. 4, Winter 2001, pp. 509–518.

be associated with accidents in the predominantly retrospective analyses reported in the literature, along with many new measures of pilot behaviors and attitudes. Using those data, I address two issues: the relationships among the various measures and accident involvement and, more broadly, the extent to which the results from retrospective studies are replicated in prospective (cohort) studies.

METHOD

During 1994 a nationwide survey of pilots was performed to collect information on their aviation qualification and experiences, participation in training activities, involvement in critical aviation incidents, personal minimums and usual practices when planning and conducting a flight, and attitudes about flying issues. Usable results were obtained from 6687 pilots. A complete description of the survey procedure, questionnaire, and comparisons of respondents and nonrespondents is provided in Hunter (1995); however, because that technical report is not widely available, the survey will be described in some detail.

Survey Data Collection

Participants. Participants were selected using simple random sampling without replacement from the population of active airman listed in the Federal Aviation Administration (FAA) Airmen Certification System. An active airman is one who has been issued a valid airman medical certificate within the preceding 25 months. The total population was approximately 561 486 pilots (excluding student pilots), from which 19 657 participants were drawn, representing private, commercial, and airline transport certificate holders.

Questionnaire development. The questionnaire was designed to provide a thorough demographic profile of the pilot population while also providing initial information on a number of areas of particular interest. These areas included training experiences, involvement in incidents that had the potential for accidents, personal preferences and practices when flying, and attitudes about flying. After a pretest with a separate sample of pilots, the final version of the questionnaire contained 143 separate items: (a) 16 questions dealing with aviation qualifications; (b) 19 questions dealing with the number of hours logged during the last 6 months, last 12 months, and during the entire career of the respondent; (c) 8 questions on the type of aircraft flown most frequently over the past year; (d) 3 questions about the careers of professional airmen; (e) 15 questions on training experiences; (f) 13 questions regarding critical aviation incidents; (g) 34 questions dealing with personal minimums and practices; (h) 27 questions concerning attitudes about flying; (i) 5 questions about participation in future research studies; and, (j) 3 questions dealing with general demographic information.

The questionnaire was printed as an optically scannable booklet and mailed to the sample of pilots along with a self-addressed business reply envelope and a cover letter explaining the nature of the study. One week after the booklets were mailed, a postcard reminder was mailed to all members of the sample.

Survey results. Of the 19 657 questionnaires mailed out, 390 were returned as undeliverable and 19 were returned because the pilot was deceased. This reduced the effective sample to 19 248. There were 6808 questionnaire booklets returned, of which 6735 were usable. The effective return rate for the survey was therefore 35%.

Of the 6735 usable booklets, 2548 were from private pilots, 2845 from commercial pilots, and 1218 from airline transport certificate holders. The associated 95% confidence intervals (for sampling error) are 2.0%, 1.9%, and 2.9%, respectively. Analyses were conducted to assess the degree of nonsampling error. This is an important consideration in a mail survey in which, as was true in the current instance, there is a large proportion of nonrespondents. If respondent or nonrespondent status is related to the purpose and content of the survey, then bias may be introduced.

To investigate this possibility, the respondent and nonrespondent groups were compared on the basis of accident rate, certificate type, gender, age, and, flight experience. No differences between the two groups were found for accident involvement, gender, or total flight time. Small but statistically significant differences were noted for certificate type, age, and recent flight time. Private pilots were slightly more likely to participate, as were older pilots and those with fewer recent flight hours. Those performing analyses that rely on those variables might need to make suitable adjustments to better estimate the population values. In the current study, the primary threat to generalizability caused by bias in the survey sample is differential accident rates, which were not found to be present.

Predictor response scales. The survey used a combination of scales, each appropriate to the data being collected. For those questions dealing with age and flight time, the response scale is the literal number of years and number of flight hours reported by the respondent. The questions that assessed involvement in hazardous aviation events had a response scale consisting of 7 scale points, labeled as 0–5, and 6+. The time scale for this set of items was the entire career of the pilot.

A 9-point scale ranging from 1 to 15 miles was used for questions concerning personal minima for visibility, and a 6-point scale ranging from 1000 to 5000 feet was used for ceiling. The responses for those questions dealing with personal procedures were scaled in terms of the percentage of time taken for the individual to perform that action, 0% to 100% broken into seven intervals, including a response of not applicable.

The opinion questions used a 5-point Likert scale (1 = *strongly agree;* 5 = *strongly disagree*). The hazardous attitude factors shown in Table 1 are dimensionless factor scores based on the aviation opinion questions that dealt with each of the five hazardous attitudes. Higher scores on these factors indicated higher agreement with statements espousing those attitudes.

Criterion Development

Because the 1994 survey was confidential but not anonymous, it was possible to associate information from both FAA and National Transportation Safety Board (NTSB) databases with the individual response records for each respondent. Thus in the analyses it was possible to evaluate the characteristics of respondents and nonrespondents to assess potential bias. It also made possible future follow-up analyses of this cohort.

In December 1999 the respondent records were matched (based on FAA certificate number) with NTSB accident records to identify pilots who had been in a reported accident either prior to the survey (presurvey) or during the approximately 5.5 years after the survey (postsurvey). In the interests of maximizing power through as large a criterion group as possible, all accidents, regardless of cause (e.g., pilot error, mechanical failure, etc.) or severity (e.g., fatal or nonfatal) were included. Involvement in any reported accident during the two periods was coded as two binary variables (presurvey and postsurvey).

Thus 182 cases were identified in the presurvey group and 89 cases in the postsurvey group. The latter number approximates the expected number of accidents (99) that would be observed based on the total number of accidents per year during the period 1994–1999 (1800 per year) and an average pilot population of 620 000.

Data Analysis

Multiple *t* tests were used to compare the means of the accident-involved pilots and the nonaccident-involved pilots for all questions in the survey. Separate analyses were conducted for the presurvey accident involvement and the postsurvey accident involvement criteria. Table 1 contains the results of the mean comparisons that achieved statistical significance at the $p < .05$ level for the presurvey analyses, and Table 2 contains the corresponding analyses for the postsurvey criterion. For each table, variables that did not reach the $p < .05$ level of statistical significant are not reported.

There were 43 comparisons that achieved significance in the presurvey analysis and 25 comparisons that achieved significance in the postsurvey analysis. In addition, 13 instances in both the presurvey and postsurvey comparisons were significant. These measures, and the effect size for both the presurvey and postsurvey analyses, are listed in Table 3. The intercorrelations of measures from Table 3 are given in Table 4.

Each set of analyses (presurvey and postsurvey) contained 157 comparisons, however, only the subsets of comparisons that achieved statistical significance are given in Table 1 or

TABLE 1: Comparison of Pilots on Accident Status Prior to Survey

	No Accident		Accident			
	Mean	SD	Mean	SD	*t*	*d*
Pilot age	49.8	13.7	52.1	12.9	–2.28*	–0.056
Flight Time – Hours						
Total time – Career	3431.5	5191	4689.5	5804.8	–2.77**	–0.07
Airplane – Career	3220.6	5154.8	4374.6	5487	–2.68**	–0.068
Single engine – 12 months	59.2	171.4	101.8	203.6	–2.66**	–0.068
Single engine – Career	1646.2	3066	2928.1	3918.7	–4.13**	–0.105
Day – Career	2673	4061.9	4008.6	5769.1	–2.83**	–0.073
Piston-powered – 6 months	28.8	61.4	43.8	73.1	–2.58*	–0.066
Piston-powered – 12 months	64.4	113.4	107.3	164.6	–3.31**	–0.085
Piston-powered – Career	1835	2836.3	3404.8	4523.5	–4.39**	–0.113
Jet – 12 months	35.8	169	18	87.3	2.36*	0.063
Instructor – Career	603.1	1773.4	965.4	2253.2	–2.00*	–0.052
Personal business – 12 months	14.3	53.6	32.8	98.7	–2.28*	–0.06
Personal business – Career	348	2081.8	676.5	1661.9	–2.40*	–0.063
Pleasure – Career	680	1001.1	1042.3	2075.6	–2.12*	–0.055
Commercial – Career	1555	4296.8	2309.8	4591.3	–2.02*	–0.053
Military – 6 months	0.96	11.3	0.027	0.33	5.97**	0.16
Military – 12 months	3.6	100.9	0.095	1.15	2.54*	0.068
Other Aviation Experience						
Landings – 12 months	110.9	244.9	196.5	473.6	–2.31*	–0.058
Landings – 6 months	48.7	146.9	91.1	237.7	–2.26*	–0.058
Practice DF approach	1.75	2.25	2.15	2.43	–2.12*	–0.053
Hazardous Events						
Self-reported aircraft accidents	0.17	0.52	1.14	0.74	–17.02**	–0.422
Off-airport precaution/forced landing	0.25	0.88	0.78	1.39	–4.91**	–0.122
Inadvertent stalls	0.14	0.65	0.33	1.11	–2.19*	–0.054
Mechanical failures	1.54	1.82	1.95	1.9	–2.77**	–0.069
Fuel starvation engine failures	0.18	0.61	0.32	0.67	–2.60**	–0.064
Turned back because of weather	2.36	2.19	2.73	2.34	–2.02*	–0.05
Made very bad decisions	1.26	1.48	1.69	1.74	–3.19**	–0.079
Personal Minima and Procedures						
Cross-country: Night minimum visibility	5.84	2	5.48	2.28	1.99*	0.051
Cross-country: Day minimum ceiling	3	1.34	2.72	1.52	2.34*	0.059
Local: Perform complete preflight	5.9	0.57	5.58	1.08	3.88**	0.097
Local: Use checklist landing & takeoff	5.4	1.46	5.02	1.75	2.76**	0.069
Cross-country: Weather brief before takeoff	5.79	0.75	5.63	0.99	2.11*	0.053
Cross-country: Complete preflight	5.96	0.45	5.8	0.8	2.59*	0.065
Cross-country: Use checklist takeoff/landing	5.48	1.4	5.22	1.64	1.99*	0.05
Cross-country: Fly VFR above clouds	1.06	1.62	0.83	1.09	2.63**	0.066
Opinions						
Am very careful pilot	0.54	0.66	0.73	0.82	–3.07**	–0.076
Rules are too strict	2.52	1.02	2.29	1.06	2.82**	0.07
Am so careful will never have accident	2.5	0.91	2.83	0.92	–4.71**	–0.117
Am very cautious pilot	0.78	0.66	0.95	0.75	–2.93**	–0.073
Unlikely I would have accident	2.4	0.91	2.73	0.92	–4.65**	–0.116
Hazardous Attitude Factors						
Caution	4.8	1.25	5.02	1.35	–2.06*	–0.052
Denial	6.85	2.11	7.57	2.1	–4.41**	–0.111
Anti-authority	2.55	1.82	2.115	1.81	2.81**	0.071

Note: $*p < .05$ (two-tailed), $**p < .01$ (two-tailed); *d* = Effect size

Table 2. Therefore, one would expect approximately 8 comparisons (5% of 157) to have been significant at the .05 level, simply by chance, if the null hypothesis of zero differences between the means were true in all cases. This could be interpreted to mean that of the 45 statistically significant findings for the presurvey analysis, 8 may have arisen by chance. However, it cannot be determined which eight those were. Typical corrections for inflation of the Type I error rate simply boost the minimum level of significance proportional to the number of comparisons conducted. Those corrections have not been applied here because a more parsimonious approach is possible that does not suppress potentially interesting results: cross-replication of the findings.

Instead of focusing on the findings from either the presurvey or postsurvey analyses in isolation (knowing that some of the results will be spurious), it is more informative to look at them in conjunction. Thus although the results shown in Tables 1 and 2 are of passing interest, the more interesting and reliable findings are in Table 3, which shows the measures that were significant in both the presurvey and the postsurvey analysis. Since the joint probability of both tests being significant is .0025 (i.e., .05 × .05) if the null hypothesis is true, then fewer than one of the 157 comparisons would be expected to be spurious (0.25% of 157 = 0.39). Thus Table 3 shows an acceptable level of protection from Type I error.

In addition to the usual *t* statistic, effect sizes for each comparison were computed using the estimation method given by Rosenthal and Rosnow (1991, p. 441), in which $d = 2t/\sqrt{df}$, and are reported in Tables 1, 2, and 3. Effect sizes are particularly interesting because, unlike *t* statistics, they may be compared to assess the relative level of association of the measures to the criterion.

RESULTS

The largest group of variables that were significantly related to accident involvement for both the presurvey and postsurvey comparisons consists of various measures of flying activity. These measures include (a) single-engine time during the last 12 months, (b) piston-powered time (for the last 6 months, last 12 months, and total career), and (c) landings in the last 6 months and last 12 months.

Previous involvement in an aircraft accident was also associated with accident involvement in the postsurvey analyses. There were two measures of previous accident involvement. One measure was self-report of the number of previous aircraft accidents obtained from the survey, which was associated significantly with postsurvey accident involvement. Not surprisingly, it was also associated significantly with presurvey accident involvement. Indeed, the failure of the association to be perfect raises questions about the reliability of the data sets, as one would expect every accident to appear in the NTSB database.

There are two obvious reasons that this is not the case. First, the NTSB database does not include accidents that occurred prior to 1983. Therefore, the self-report measure may include accidents that predated the NTSB database. Second, pilots may use a more liberal definition of accident than that which determines entry into the NTSB database. For example, a fairly minor incident involving damage to the aircraft might be considered an accident to the pilot (particularly if he or she had to pay for the repairs), but it would not meet the definition of the FAA or NTSB and hence would not appear in the database.

For these reasons, the accident involvement measure obtained from the NTSB database was considered a more reliable measure of accident involvement. This measure (Presurvey Accident Involvement, listed last in Table 1) indicated whether the pilot had been involved in an accident that had been reported to the NTSB prior to the survey and was significantly ($t = -2.079$, $p < .05$) associated with postsurvey accident involvement.

Off-airport precautionary or forced landings were also associated with accident involvement in both the presurvey and postsurvey analyses. This is most likely a surrogate measure for accident involvement – that is, usually an off-airport landing is, per se, an accident. However, it might also reflect exposure to potential accidents.

The remaining measures all seem to involve the pilots' perceived and actual level of caution.

TABLE 2: Comparison of Pilots on Accident Status following Survey

	No Accident		Accident			
	Mean	SD	Mean	SD	*t*	*d*
Flight Time – Hours						
Total time – 12 months	127.3	228	183.4	226.5	–2.28*	–0.057
Single engine – 12 months	59.9	173.1	86	112.1	–2.08*	–0.053
Day – 12 months	106.1	236	150.8	154.2	–2.48*	–0.064
Simulator – 12 months	3.9	27.9	2	6.2	2.41*	0.063
Simulator – Career	77.7	399.1	39.8	91.8	3.30**	0.085
Piston-powered – 6 months	28.9	61.4	50.2	84.2	–2.30*	–0.059
Piston-powered – 12 months	65	115.1	101.2	118.3	–2.77**	–0.071
Piston-powered – Career	1865.7	2897.3	2720.6	3706.7	–2.07*	–0.053
Pleasure – 12 months	32.9	132.2	47.9	46.6	–2.70**	–0.07
Other Aviation Experience						
Landings – 12 months	111.1	249.1	271	475.2	–3.02**	–0.076
Landings – 6 months	50	148.8	113.6	222.5	–2.58*	–0.066
Instrument approaches – 6 months	18.3	38.4	16	20.7	2.06*	0.052
Hazardous Events						
Self-reported aircraft accidents	0.19	0.54	0.45	0.85	–2.84**	–0.07
Off-airport precaution/forced landing	0.26	0.9	0.51	1.13	–2.03*	–0.05
Flew VFR into IMC	0.35	0.87	0.55	0.88	–2.03*	–0.05
Personal Minima and Procedures						
Local: Day minimum visibility	2.68	1.5	2.31	1.6	2.09*	0.052
Local: Day minimum ceiling	1.53	1.23	1.14	1.22	2.98**	0.075
Cross-country: Day minimum ceiling	3	1.35	2.57	1.43	2.75**	0.069
Local: Use checklist landing & takeoff	5.39	1.46	4.88	2.05	2.31*	0.058
Cross-country: Use checklist takeoff/landing	5.48	1.4	4.95	2	2.43*	0.061
Opinions						
Am very careful pilot	0.54	0.66	0.72	0.64	–2.55*	–0.063
Am very cautious pilot	0.78	0.66	0.99	0.73	–2.64**	–0.066
Fly enough to maintain proficiency	1.53	1.12	1.06	0.88	4.94**	0.123
Hazardous Attitude Factors						
Impulsive	3.8	1.36	4.1	1.3	–2.05*	–0.051
Presurvey accident involvement	0.026	0.16	0.09	0.29	–2.08*	–0.051

Note: * $p < .05$ (two-tailed), ** $p < .01$ (two-tailed); d = Effect size

Three measures deal with behaviors reflective of caution: (a) cross-country minimum day ceiling, (b) using a checklist for landing and takeoff during local flights, and (c) using a checklist for landing and takeoff during cross-country flights. All three are positively related to safety – that is, pilots who use checklists or have more conservative ceiling minima are less likely to be involved in an accident. The two self-assessment items were also related positively to safety. For those two items ("I am a very careful pilot"; "I am a very cautious pilot"), pilots who agreed with the statements were less likely to be involved in an accident.

CONCLUSIONS

Most of the measures that were found to be significant statistically may be considered measures of exposure. These include simple measures such as various aspects of total time, which operated primarily in retrospect, and measures of recent flying activity, which are probably a reasonable measure of expected future activity levels. It is reasonable to assume that the more a pilot flies, the more often he or she is put at risk of an accident. However, without flying, a pilot cannot gain the experience that is usually regarded as a protective factor. Exposure mea-

TABLE 3: Measures Found To Be Significantly Associated with Accident Status in Both Presurvey and Postsurvey Analyses

	Effect Size	
	Presurvey	Postsurvey
Flight Time – Hours		
Single engine – 12 months	–0.068	–0.053
Piston-powered – 6 months	–0.066	–0.059
Piston-powered – 12 months	–0.085	–0.071
Piston-powered – Career	–0.113	–0.053
Other Aviation Experience		
Landings – 12 months	–0.058	–0.076
Landings – 6 months	–0.058	–0.066
Hazardous Events		
Self-reported aircraft accidents	–0.422	–0.07
Off-airport precaution/forced landing	–0.122	–0.05
Personal Minima and Standards		
Cross-country: Day minimum ceiling	0.059	0.069
Local: Use checklist landing & takeoff	0.069	0.058
Cross-country: Use a checklist takeoff/landing	0.05	0.061
Opinions		
Am very careful pilot	–0.076	–0.063
Am very cautious pilot	–0.073	–0.066

sures are a sterile area for those interested in understanding the causes of accidents and the development of interventions to reduce accident risk.

Baker et al. (1993) examined NTSB records of crashes of commuters and air taxis during the period 1983–1988. They found that "Pilots in crashes were almost three times as likely as comparison pilots to have had an aircraft accident in the prior three years…[and]…Pilots in crashes were twice as likely as comparison pilots to have an accident in the next three years" (p. 4). Similar results were noted by Li and Baker (1994) in a case-control study of commuter and air taxi crashes. The results from the current study support those previous findings. Analysis of both the self-report data obtained from the survey and the NTSB official records suggest that having been in an accident increases the risk of future accidents.

This is simultaneously intuitively obvious and profoundly disheartening. From the standpoint of behavioral consistency, it should be obvious that the environment, habits, practices, and behaviors that contributed to one accident are unlikely to change and, hence, continue to place a pilot at increased accident risk. At the same time, this is truly depressing news to those charged with developing interventions to change pilot habits, practices, and behaviors so as to improve aviation safety. Surely, one might think, having been in an accident would provide a tremendous stimulus for positive change. It appears not to do so, so motivating pilots to change when one can only describe what *might* happen to them and how training *might* prevent it will be a tremendous challenge.

The finding that pilots who consistently use a checklist during both local and cross-country flights are less at risk for accident involvement could be attributed to, at least, two potential causes. This might be interpreted as a general indication of pilot caution as reflected in the pilot's attention to detail. Alternatively, it could be argued that simply following the checklist, in and of itself, will reduce accident risk because fewer lapses and procedural errors will occur. One interpretation takes the observed relationship as an indicator of a deeper personality aspect, whereas the other interpretation is simply that doing safe things results in safety.

TABLE 4: Intercorrelation of Measures Significantly Associated with Accident Status in Both Presurvey and Postsurvey Analyses

	Hours: SE 12 mo.	Hours: Piston 6 mo.	Hours: Piston 12 mo.	Hours: Piston Career	Landing 12 mo.	Landing 6 mo.	Self Report Accident	Off-Airport Landing	Minimum Day Ceiling	Check-list: Local	Checklist: Cross-Country	Am Careful Pilot
Hours: Single Engine – last 12 mo.	1											
Hours: Piston Power – last 6 mo.	.515**	1										
Hours: Piston power – last 12 mo.	.606**	.791**	1									
Hours: Piston Power – Career	.123**	.262**	.324**	1								
Landings: Number in last 12 mo.	.202**	.257**	.353**	.176**	1							
Landings: Number in last 6 mo.	.151**	.208**	.246**	.126**	.724**	1						
Number of Self-report accidents	.032*	.041**	.042**	.241**	.026*	.037**	1					
Number of off-airport landings	.051**	.016	.025	.211**	.124**	.076**	.241**	1				
Minimum Cross-Country Day Ceiling	–.042**	–.097**	–.105**	–.176**	–.092**	–.113**	–.108**	–.149**	1			
Use Checklist: Local	–.047**	–.092**	–.090**	–.093**	–.065**	–.037**	–.080**	–.064**	.136**	1		
Use Checklist: Cross-Country	–.083**	–.083**	–.089**	–.096**	–.073**	–.038**	–.063**	–.048**	.136**	.918**	1	
Am a very careful pilot	.023	.014	.025	.018	–.003	.011	.089**	.052**	–.005	–.027*	–.024	1
Am a very cautious pilot	.026*	.012	.018	.052**	.001	–.005	.095**	.063**	–.088**	–.157**	–.137**	.206**

Note: * $p < .05$ (two-tailed), ** $p < .01$ (two-tailed); d = Effect size

The finding that pilots who believe themselves to be careful and cautious have a lower risk of accident involvement is intriguing. Earlier studies (Hunter, 1995, 1997) suggest that pilots have an exaggerated sense of their personal capabilities, which may lead them to venture into situations beyond their skill level. However, the present results are clearly inconsistent with that interpretation. The pilots in the present study seemed to have an accurate perception of their relative degree of carefulness and caution, given that those who rated themselves most highly on those traits had the fewest accidents. Conversely, those pilots who rated themselves lower on those traits (even though almost none rated themselves at the bottom of the scales) had more accidents.

One can only wonder why pilots who, with apparent accuracy, believe themselves to be less than completely rigorous in two such important matters do not take actions to control these factors and lower their risk of an accident. Do they not perceive a relationship between their actions and the risk of an accident? Are they aware of the relationship but tolerant of the increased risk? The relationships among these factors are probably complex and in need of additional research, as they speak to the heart of potential intervention strategies.

Contrary to Lategola et al. (1970), the present study did not find prospective validity for age. However, as Urban (1984) found, there was a trend for advanced degree holders (i.e., M.D., J.D. Ph.D.) to have a higher accident rate. Booze (1977) also noted a higher accident rate among lawyers and physicians and attributed this to common personality traits of aggression, independence, and self-sufficiency. However, commenting on the Booze results, O'Hare (1999) suggested that "rather than sharing a raft of personality characteristics, it is more likely that lawyers and physicians have access to more complex, higher performance aircraft…thus exposing themselves to an increased variety of operational risks" (p. 268).

With respect to the second study objective, the analyses indicated that most variables that showed significant relationships with accident status in the retrospective study failed to achieve significance in the prospective (cohort) study. Specifically, of 43 variables that were significant in the retrospective analysis, only 13 (29%) were significant in the prospective analysis. One explanation for this difference is that the second analysis used a considerably smaller criterion group and hence did not have the statistical power to detect true effects of the small magnitude observed in the first analysis. We may evaluate this explanation by observing the effect sizes obtained in the two analyses, given that these estimates are independent of the sample size. Examination of the effect sizes shown in Tables 1 and 2 indicates that, for the most part, they do not support the limited power explanation.

Of the 30 mean comparisons that were significant in the presurvey analysis but that failed to be replicated in the postsurvey analysis, none of the effect sizes for the postsurvey analysis was greater in magnitude than the effect sizes for the presurvey analysis. With only three exceptions, the effect size for the presurvey analysis was considerably larger than the corresponding effect size for the postsurvey analysis. Therefore, the failure to find statistical significance in the replication (postsurvey) analysis was not attributable to a lack of a sufficient sample but rather to an underlying lack of effect on accident involvement by the variables under examination.

Three possible exceptions to this finding were noted for the variables: (a) personal business – 12 months, (b) pleasure – career, and (c) turned back due to weather. Although the effect size in the postsurvey analysis was smaller than the effect size for the presurvey analysis for all these variables, the differences were small, suggesting that some real underlying difference may exist. In these cases, lack of sufficient statistical power may have prevented the detection of a real effect.

Overall, however, it seems clear that some measures found to be significant predictors in retrospective analyses have a very high likelihood of not predicting future accident involvement. Given that that is the very criterion of interest (predicting past accidents is, after all, rather easy if one has the NTSB database), the data raise some substantial concerns over the utility of retrospective analyses for aircraft accident research. Measures that are correlated with past performance do not necessarily predict future performance.

In summary, this study used a unique set of data available from a cohort of pilots who completed a survey seven years ago, both to examine the validity of the biodata and other personal characteristics and to evaluate the use of retrospective analyses in aircraft accident research. The results illustrate that exposure, when measured any of a number of ways, is associated with accident involvement. This is not a unique revelation, but it should serve to remind other researchers that failure to account for exposure when conducting aviation accident research seriously weakens their studies. More important, although case control and similar techniques have many advantages and are often the only analytic means available, the results show the need for caution in interpreting the results of retrospective analyses. Many of the measures examined here that showed significant associations with accident involvement in the retrospective analysis failed to show prospective validity.

REFERENCES

Baker, S. P., & Lamb, M. W., Li, G.H., & Dodd, R.S. (1993). Human factors in crashes of commuter airplanes and air taxis. *Aviation, Space, and Environmental Medicine, 64,* 63–68.

Booze, C. F. (1977). Epidemiologic investigation of occupation, age, and exposure in general aviation accidents. *Aviation, Space, and Environmental Medicine, 48,* 1081–1091.

Harkey, J. A. Y. (1997). Age-related changes in selected status variables in general aviation pilots. *Transportation Research Record, 1517,* 37–43.

Hunter, D. R. (1995). *Airman research questionnaire: Methodology and overall results* (Report DOT/FAA/AM-95/27). Washington, DC: Federal Aviation Administration.

Hunter, D. R. (1997). *An evaluation of safety seminars* (Report DOT/FAA/AM-97/16). Washington, DC: Federal Aviation Administration.

Lategola, M. T., Fiorica, V., Booze, C. F., & Folk, E. D. (1970). *Comparison of status variables among accident and non-accident airmen from the active airman population* (Report AM-70-18A). Washington, DC: Federal Aviation Administration.

Li, G. (1994). Pilot-related factors in aircraft crashes: A review of epidemiologic studies. *Aviation, Space, and Environmental Medicine, 65,* 944–952.

Li, G., & Baker, S. P. (1994). Prior crash and violation records of pilots in commuter and air taxi crashes: A case-control study. *Aviation, Space, and Environmental Medicine, 65,* 979–985.

Lubner, M. E., Markowitz, J. S., & Isherwood, D. A. (1991). Rates and risk factors for accidents and incidents versus violations for U.S. airmen. *International Journal of Aviation Psychology, 1,* 231–243.

McFadden, K. L. (1997). Predicting pilot-error incidents of U.S. airline pilots using logistic regression. *Applied Ergonomics, 28,* 209–212.

O'Hare, D. (1999). Safety is more than accident prevention. In D. O'Hare (Ed.), *Human performance in general aviation,* 265–290. Brookfield, VT: Ashgate.

O'Hare, D., Chalmers, D., & Bagnall, P. (1996). *A preliminary study of risk factors for fatal and non-fatal injuries in New Zealand aircraft accidents.* Final report to the Civil Aviation Authority of New Zealand. Dunedin, New Zealand: Department of Psychology, University of Otago.

Platenius, P. H., & Wilde, G. J. S. (1989). Personal characteristics related to accident histories of Canadian pilots. *Aviation, Space, and Environmental Medicine, 60,* 42–45.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Urban, R. F. (1984). Comparative analysis of social, demographic, and flight-related attributes begween accident and nonaccident general aviation pilots. *Aviation, Space and Environmental Medicine, 55,* 308–312.

David R. Hunter received his Ph.D. in educational psychology from the University of Texas at Austin in 1982. He is a research psychologist with the Office of Aviation Medicine of the Federal Aviation Administration.

Date received: October 4, 2000
Date accepted: January 8, 2001