Federal Aviation Administration

DOT/FAA/AM-13/4
Office of Aerospace Medicine
Washington, DC 20591

Effects of Training School Type and Examiner Type on General Aviation Flight Safety

William Knecht
Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

Jeffrey Smith
Federal Aviation Administration
Aviation Safety–Flight Standards Service
Washington, DC 20003

March 2012

Final Report

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-13/4		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Effects of Training School Type and Examiner Type on General Aviation Flight Safety	5. Report Date: March 2013 6. Performing Organization Code:
7. Author(s)	8. Performing Organization Report No.
Knecht WR,[1] Smith J[2]	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS) / 11. Contract or Grant No.
[1]FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125 [2]FAA Aviation Safety Flight Standards Service Washington, DC 20003	
12. Sponsoring Agency name and Address	13. Type of Report and Period Covered / 14. Sponsoring Agency Code
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	

15. Supplemental Notes

Work was accomplished under approved task AHRR521

16. Abstract

This study addresses the question "Do *training school type* and *certifying examiner type* affect a general aviation pilot's subsequent aviation safety record?" "Education" was operationalized as *private pilot instruction* in either a

- Part 61 or
- Part 141 school

and "examiner type" was operationalized as *private pilots examined* by either

- Aviation Safety Inspector (ASI),
- School Authority (Part 141 graduates only), or
- Designated Pilot Examiner (DPE)

Because of the unavailability of earlier reliable FAA school and examiner records, results herein are restricted to pilots certificated from 1 Jan., 1995 to 8 Aug., 2007.

The results essentially imply that that school and examiner type do not affect subsequent accident rate. For U.S. GA pilots receiving the private pilot certificate from 1995-2007 and for whom data could be obtained—Part 61 graduates' subsequent accident rate appeared on a par with Part 141 graduates, and pilots tested by DPEs appeared equivalent to those tested under school authority. Graduates tested by ASIs showed a statistically lower accident rate, but that particular result was based on a sample of only 22 pilots, rendering it unreliable from a practical point of view.

Recommendations include a) adoption of a common pilot identification number ("UniqueID") for both FAA and NTSB, to minimize data loss, and b) that user's manuals be made publically available for FAA's CAIS and DIWS databases.

17. Key Words	18. Distribution Statement
General Aviation, Training, Examination, Accident Frequency, Accident Rates	Document is available to the public through the Internet: www.faa.gov/go/oamtechreports

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	31	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

The authors would like to extend special thanks to Joe Mooney, FAA (AVP-210), for his help with the NTSB queries, to Carla Colwell and Tona Gates, FAA (AFS-760), for their help with the CAIS queries, and to David Nelms, FAA (AAM-300), for his help with the DIWS queries. Without the dedicated professionalism of these individuals a project of this complexity, extending across three separate databases, would have been impossible.

We also extend deepest thanks to Loren Groff, Ph.D., NTSB, for clarifying search characteristics of the NTSB database, to Dana Broach, Ph.D., FAA (AAM-500) for statistical consultation on log-linear analysis, and to Elaine Pfleiderer, FAA (AAM-500) for help with SPSS data-preparation issues.

CONTENTS

INTRODUCTION

In 2005, the U.S. National Transportation Safety Board (NTSB) released Safety Study NTSB/SS-05-01 *Risk factors associated with weather-related general aviation accidents*. This study included a number of recommendations, including A-05-027:

> Develop a means to identify pilots whose overall performance history indicates that they are at future risk of accident involvement, and develop a program to reduce risk for those pilots.

This recommendation was largely addressed by the FAA. However, two questions remained concerning what effects general aviation (GA) pilots' type of *education* and *certification testing* might have on their subsequent flight safety record. We attempt to address these questions in the current study.

METHODOLOGY

Research Hypothesis

Following standard procedure, we begin with the null hypotheses that pilots'

- Type of education
- Type of certification testing

have no significant effect on their subsequent flight safety record. We then design a research method to test available data to statistically confirm or disconfirm these hypotheses.

Basic Research Design

Operationalizing "flight safety record." "Flight safety" can be measured in various ways. Li (1994) noted that aviation-risk studies usually examine some sort of quotient based on

$$\frac{\textit{Frequency of some event}}{\textit{Some estimate of risk exposure}}$$

For instance, this quotient may be accidents per year or accidents per 100,000 flight hours. In the current study, we operationalize "flight safety record" as "accidents per unit time," with the "unit" defined as a time period spanning several years, to capture a greater number of events.[1]

Operationalizing type of education and type of certification testing. "Type of education" covers too broad a swath to be investigated fully, given the many types of pilot certificates. Therefore, based on the logic that the *private pilot certificate* is universal among the vast majority of pilots and may indeed be the only certificate a GA pilot ever gets, we first operationalize "education type" as *whether a GA pilot received the private pilot certificate from a Part 61*[2] *versus a Part 141 school*, as defined in Title 14 of the U.S. Code of Federal Regulations, Title 14, Part 67 (§67.121.309(d)). This does exclude recreational and sport pilots, however; these constituted less than 1.5% of all new airman certificates issued during the time period studied (FAA, 2010). Further, it should be noted that it is possible for a student to have received initial private pilot training from both a part 61 and a 141 pilot school, and that the pilot classification in this study would refer to the regulatory part the pilot was certificated under.[3] It is reasonable to conclude that the regulatory part the pilot was certificated under made the final assessment as to the pilot's proficiency and ability to pass the private pilot practical test. It is this final regulatory part that will be associated with the pilot's education type. Operationally, we shall label this variable "School type" (abbreviated as "School").

Similarly, we can operationalize "certification testing" as whether that pilot's private certificate examiner was an Aviation Safety Inspector (ASI), personnel of a flight school that holds examining authority (BYSCHOOL, Part 141 only), or Designated Pilot Examiner (DPE). We label this variable "Examiner type" (abbreviated as "Examiner").

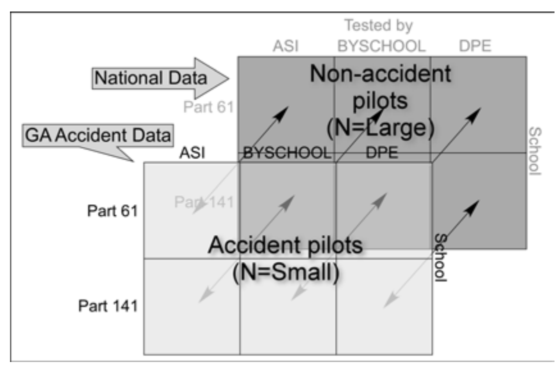


Figure 1. The basic 2x2x3 analytical design.

The basic analytical design. The operational definitions just given suggest a basic structure for an analytical design, shown in Figure 1.

To calculate the statistical effects of School and Examiner on subsequent accidents, we essentially need to compare GA accident pilots against some standard or baseline. For instance, accident pilots can be compared to non-accident pilots.

[1] We fully realize that some readers will be disappointed and would prefer to see a study based on, say, accidents per flight hour, or per departure. Unfortunately, that kind of information is simply not readily available at this time.

[2] Technically, there are no "Part 61 schools. Rather, there are flight instructors, or collections of flight instructors, operating under the Part 61 authority of their individual certificates, instead of the formal authority granted to an actual flight school (as under Part 141). However, since it is colloquial and useful to call these "Part 61 schools," we follow that convention here.

[3] As in the previous footnote, we acknowledge that all private pilots are technically certificated under Part 61 (even if they graduated from a Part 141 school). But, since it is "the effect of school" we are after here, we again choose to speak colloquially.

In this type of design, each of the 12 cells in the 2x2x3 "Accident x School x Examiner" matrix contains the number of individuals—the *frequency count*—in that set of conditions. The front "Accident matrix" contains the numbers of pilots who subsequently had an accident after receiving their private pilot certificate during the time period under examination, with pilots assigned to cells by the type of school they attended and the type of examiner that tested them. Similarly, the rear "Non-accident matrix" consists of a large national group of non-accident pilots parsed the same way, by rows and columns. We expect that the non-accident pilots will greatly outnumber the accident pilots. And, foremost in our minds will be determining the relation between School, Examiner, and subsequent pilot accidents.

The Data

Adding school and examiner type. Initially, FAA's Office of Accident Investigation and Prevention (AVP-210) provided a list of all pilots involved in *serious-injury or fatal GA accidents* taken from NTSB's database, encompassing a time period from Jan. 1, 2003 to Aug. 28, 2007 (4 yr, 8 mo, N_{pilots}=7,342).[4] "GA aircraft" were defined as "all N-tail-numbered aircraft operating under all Federal aviation regulations Parts except 121 and 135, regardless of airframe type or weight."[5] Although varying distinctions between "general aviation" and "non-general aviation" could be argued, defining general aviation in this manner is consistent with FAA precedent and provides a reasonable grouping for the purposes of our study. Those 7,342 cases were next given to FAA's Office of Flight Safety (AFS-760), whose staff were able to match 2,090 of those accident pilots (28.5%) to the data mandated by our study, namely:

- School Type   Part 61 vs. Part 141 schools (with Part 142 classified as 141).[6]
- Examiner Type   Aviation Safety Inspector (ASI) versus Tested By School Authority (Part 141 schools only) versus Designated Pilot Examiner (DPE).

This matching was done by cross-referencing listed NTSB pilot certification numbers and/or names to the FAA Comprehensive Airman Information System database (CAIS, pronounced "CASS"), which contains school and examiner information. The low match rate was due to a number of reasons: a) the NTSB pilot certificate field (labeled "crew_cert_id") did not match the CAIS pilot certificate field, making retrieval of that pilot's school and examiner data impossible,[7] or because; b) CAIS contains school and examiner data only from 1995 on. Since many of our accident pilots had received their first certificate before that, their school and examiner data were therefore missing. This constraint had statistical ramifications, which will be discussed wherever appropriate.

Additional exclusions. Additional pilots had to be excluded for a variety of reasons. For instance, 12 foreign-national pilots were excluded because their CAIS data reflected dates when they first received a certificate in the U.S., so training received in their native country was unrecorded. Fourteen additional pilots were excluded because the date of their accident was listed as being prior to the date of their private pilot certification. This would be consistent with having an accident while still being a student (however, this was unknown; these could have been data-entry errors). More importantly, we were interested in how school and examiner type might subsequently affect accidents after graduation; therefore, students who had not yet earned their private pilot certificate were not a group of interest here.

Next, we attempted to exclude all persons other than pilots-in-command (PIC).[8] NTSB data list all persons involved in an accident, regardless of whether they actually exerted any influence on how that accident unfolded. Since we were primarily interested in the person most likely to have been able to prevent each accident, we chose to focus on PICs, while excluding all others. Our parsing method was based on the logic that the PIC should be the senior pilot onboard, ultimately in control of the aircraft facing an impending accident,[9] and therefore the person of interest when determining how school and examiner type might affect accidents.

In practice, identifying the PIC can be difficult. In non-fatal accidents, on-scene investigators can interview the flight crew. But, if one can only look at a row of data from a database, the PIC may be ambiguous when multiple persons are onboard and/or when there are no survivors. In point of fact, the NTSB does identify a field in their database designated as PIC.[10] However, in practice, that person is typically assumed to be the pilot identified at the controls.[11] In cases such as flight instruction,

[4] NTSB only infrequently grants FAA a limited number of full copies of its database (having pilot names and certification numbers). Those were necessary to match each pilot with his/her specific flight school and examiner type. Therefore, we were limited to using the most-current NTSB database available, which ran to Aug, 2007.

[5] Our data contained no Part 129 pilots (foreign air carriers operating N-registered aircraft).

[6] The Part 142 (§142) training centers were nominally grouped with Part 141 schools as both entity types provide instruction under approved training programs. In actuality, our data contained no Part 142 pilots listed as such.

[7] Older pilots used to have certificate numbers matching their 9-digit Social Security numbers. As of June 2002, that policy was changed for privacy reasons, and 7-digit certificate numbers began to be issued. CAIS contains whichever number each pilot prefers. However, this can result in mismatch with NTSB's record for the same pilot.

[8] The NTSB data dictionary included with their database defines PIC as: "Pilot or pilot-in-command means the person who 1) has final authority and responsibility for the operation and safety of the flight, 2) has been designated as pilot-in-command before or during the flight, and 3) holds the appropriate category, class, and type rating, if appropriate, for the conduct of the flight. Title 14 CFR, 91.3 designates the pilot-in-command of an aircraft as being directly responsible for and the final authority as to the operation of that aircraft. In general, 14 CFR, 61 prescribes certification requirements to act as pilot-in-command of various flight operations."

[9] Note that we are not implying that the PICs "caused" their accident—only that they were defined as PIC by the selection standards imposed here in order to conduct the present study.

[10] NTSB Table *Flight_Crew*, field *crew_category*, designator *PLT*.

[11] Personal communication, L. Groff, Ph.D., NTSB, January 19, 2011.

Table 1. Accident data grouped by school, examiner, and instrument rating.

Examiner				School	Examiner				Instrument-rated at time of accident?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
16	0	1,677	*1,693*	Part 61	6	0	974	*980*	No
6	78	91	*175*	Part 141	1	19	36	*56*	
22	*78*	*1,768*	*1,868*	Total	*7*	*19*	*1,010*	*1,036*	
Whole group				Part 61	10	0	703	*713*	Yes
				Part 141	5	59	55	*119*	
				Total	*15*	*59*	*758*	*832*	

a valid argument can be made that the flight instructor, being the most experienced pilot aboard, should be held "statistically accountable" for purposes such as ours.

In our accident data, the inclusion/exclusion process was simple for single-pilot accidents—all were included. However, the process was more involved for cases of multiple-pilot and multiple-aircraft accidents. The NTSB case numbers listed all pilots and aircraft in any given accident together under the same case number. So, we ultimately tried to determine which pilots were in which planes, and then assign a "status hierarchy" to each crewmember to determine who should be delegated as PIC in each aircraft.

Below is the initial crew status categorization[12] assigned by us to the original 7,342 cases. These are roughly rank-ordered by "command status," defined as "degree of command authority, given the type of flight":

1. Flight Instructor .... *n*=613
2. Check Pilot 33
3. Pilot 6,281
4. Co-pilot 149
5. Flight Engineer......................... 2
6. Student................................ 175
7. Other 45
8. Unknown (blank data cell) 44

In most cases, this hierarchy was sufficient to determine PIC and led to additional pilots being excluded as non-PICs. For instance, all co-pilots, flight engineers, students, "other," and "unknown" were excluded.

A total of 115 residual cases resisted simple determination of PIC. We therefore manually checked those individual, associated NTSB accident reports themselves. While laborious, this method increased the chance that only PICs-as-defined were included in the final accident data file. The net result was an additional 75 of the residual 115 pilots being excluded for not being PICs.

Determining pilot instrument rating and total flight hours at time of accident. While not part of the original FAA response to the NTSB, two additional factors of interest are pilot instrument rating and flight risk associated with total flight hours (TFH). The risk factors faced by instrument-rated (IR) pilots are arguably considerably different than those faced by non-instrument-rated (NIR) pilots. Likewise, we know that most GA accidents happen to relatively low-hour pilots (Craig, 2001). Moreover, it is always wise during statistical analysis to make some attempt to control for risk exposure; for instance, by considering TFH as a covariate.

Our original records from NTSB did not state whether each pilot was IR at the time of their accident. But, it seemed logical to examine instrument rating as a potential factor possibly distinguishing accident pilots from non-accident pilots. So, we first tried deriving it by comparing accident dates (from NTSB) to IR issuance dates (from CAIS). If the CAIS IR issuance date preceded the accident date, the pilot was judged "IR at the time of accident."

However, it soon became evident that this method was flawed. To properly declare a given pilot "IR at the time of accident," that pilot should not just hold *any* type of instrument rating, but rather the type of rating appropriate to the type of aircraft involved.[13]

We therefore requested the NTSB's record of which pilots were IR at the time of accident. Generally, CAIS and NTSB records agreed, but not completely (94%). So, we decided to accept the NTSB record as the standard, since these involved an investigator present on-scene who also later obtained the pilots' ratings from FAA records.[14]

Net result. After all the above mentioned exclusions, the net data-survival was 1,868 pilots identifiable as being PICs having school data, examiner data, plus IR data (25.4% of the original 7,342). These cases spanned a final accident-data range of Jan 1, 2003 to Aug 26, 2007 (a net time loss of just two days from the original date range).

Table 1 shows these 1,868 pilots grouped by School, Examiner, and Instrument Rating. Whole-group data are at left of the table; to the right, are data parsed by instrument rating. According to NTSB records, 1,036 pilots (55.5%) were non-instrument-rated (non-IR) at the time of the accident. This approximates the relative percentages found in the pilot

[12]The actual NTSB data field is called *crew_category*.

[13]There are several types of airframe categories—airplane, helicopter, glider, gyrocopter, balloon, powered lift, and blimp (in descending order of frequency). A powered lift is a rare category of aircraft such as the Harrier jet or tilt-rotor Osprey. However, at the time for which we had data, the FAA granted instrument ratings only for airplanes, helicopters, and powered lifts.

[14]NTSB reportedly requests "Blue Ribbon Packages" from FAA—individual, comprehensive records for each airman involved in an accident. These contain all the airman's instrument ratings, with dates of issuance (L. Groff, NTSB, personal communication).

Table 2. Non-accident data grouped by school, examiner, and instrument rating.									
Examiner				School	Examiner				Instrument-rated?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
462	0	55,104	55,566	Part 61	291	0	34,088	34,379	No
253	3,649	6,351	10,253	Part 141	122	1,315	2,475	3,912	
715	3,649	61,455	65,819	Total	413	1,315	36,563	38,291	
Whole group				Part 61	171	0	21,016	21,187	Yes
				Part 141	131	2,334	3,876	6,341	
				Total	302	,2334	24,892	27,528	

Table 3. Adjusted non-accident data (Table 2 – Table 1).									
Examiner				School	Examiner				Instrument-rated?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
446	0	53,427	53,873	Part 61	285	0	33,114	33,399	No
247	3,571	6,260	10,078	Part 141	121	1,296	2,439	3,856	
693	3,571	59,687	63,951	Total	406	1,296	35,553	37,255	
Whole group				Part 61	161	0	20,313	20,474	Yes
				Part 141	126	2,275	3,821	6222	
				Total	287	2,275	24,134	26,696	

population at large (59.1% IR).[15] The remaining 832 pilots (44.5%) were instrument-rated (IR) in the category of aircraft being operated at the time of the accident.

The national "non-accident" group. As Figure 1 previously illustrated, the basic analysis called for comparing our accident data to a large national group of non-accident pilots to look for differences. To that end, AFS-760 also provided nationwide CAIS data for 302,685 pilots, as a "snapshot" and containing the same key information as our accident data—particularly, private pilot a) school, b) examiner, c) issuance date,[16] and d) instrument rating.[17]

We then sent the data to AAM-300 to add most-recent total flight hour (TFH) estimates reported during pilot medical certification. For equilibration purposes, TFH data were constrained to lie within the same time window as the accident data.[18] After initial difficulty, we were told of a publically undocumented "UniqueID" number shared by both FAA databases,[19] which enabled better matching between CAIS and DIWS. However, after adding the constraint of first issuance date,[20] usable data containing both first issuance *and* TFH were severely restricted to about 23%. After finally removing all pilots with < 45 TFH (assumed to be students) and removing all pilots with more than 65,000 TFH (assumed to be either reporting errors or data-entry errors),[21] a net 65,819 cases (21.7%) remained. Table 2 summarizes.

Assuming that these 65,819 cases *also contained our* 1,868 *accident pilots* (2.8%), we subtracted out those 1,868 accident pilots, to leave a purer, "non-accident" group, against which to compare our accident data. The easiest way to do this was to simply subtract Table 1 from Table 2, to produce Table 3.

Total flight hours. Table 4 shows the aggregated (summed) total flight hours for accident pilots (top table) and non-accident pilots (bottom table). For non-accident pilots, the accident TFH were subtracted from the raw data to produce the adjusted TFH shown.

Potential Biases in the Data Inclusion/Exclusion Process

At this point, it is appropriate to raise the issue of whether any significant, systematic biases may have occurred in selecting the data. Such biases could affect our conclusions. Unfortunately, biases are easy to introduce when excluding data that fail to meet some selection criteria or when an original dataset itself contains some inherent bias. It can sometimes be very hard to distinguish

[15]Source: www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2007/media/07-air4.xls, averaged over years 1998-2007, defined as $N_{\text{IR pilots}} / (N_{\text{all pilots}} - N_{\text{student pilots}})$.

[16]Table 2's basis for determining instrument rating involved simply whether or not pilots possessed any instrument rating at the time we requested this sample. As such, it is a "snapshot" of the GA population.

[17]Foreign-certificated pilots were absent from this group, since they had no school or examiner-type data entered into CAIS.

[18]For U.S. non-accident pilots, the best available flight hour estimates currently come from the FAA's Aerospace Medical Certification Division/Document Imaging Workflow System (AMCD/DIWS, AAM-300), transcribed from FAA Form 8500-8 gathered during pilot medical re-certification. For the private GA pilot, this involves Class-3 medical certification, recurring every 5 years for pilots under 40 years of age, and every 2 years thereafter.

[19]Neither CAIS nor DIWS has a publically available user's manual.

[20]CAIS contains information on all U.S. pilots—but not all information is reliably present for those who first became pilots before 1995. If a private-pilot issuance date is present, then school, examiner, and IR issuance date will also be reliably present. Therefore, issuance date became a filter criterion to better equilibrate our accident and national samples. This eliminated potential uncontrolled biases due to inconsistent data collection for older pilots, though at the cost of the data now only reflecting pilots private-certificated after Jan 1, 1995.

[21]The 65,000 TFH cutoff was arbitrary but as liberal as any logical person could defend. Assuming a pilot flies 6 hours/day, 6 days/week, 50 weeks/year, a TFH of even 65,000 equals 36.1 years of flying. The odds of any Figure greater than that being accurate seem remote.

Table 4. Aggregate pilot total flight hours, accident data (term FH_{ijkl} in Equation 1 below).

Examiner				School	Examiner				IR?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
27,081	0	1,301,766	*1,328,847*	Pt 61	5,837	0	463,067	*468,904*	No
4,031	73,284	79,637	*156,952*	Pt 141	105	9,006	13,355	*22,466*	
31,112	*73,284*	*1,381,403*	*1,485,799*	Total	*5,942*	*9,006*	*476,422*	*491,370*	
Whole group				Pt 61	21,244	0	838,699	*859,943*	Yes
				Pt 141	3,926	64,278	66,282	*134,486*	
				Total	*25,170*	*64,278*	*904,981*	*994,429*	
Adjusted non-accident TFH (raw TFH - accident TFH).									
Examiner				School	Examiner				IR?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
637,392	0	41,524,050	*42,161,442*	Pt 61	325,086	0	13,358,621	*13,683,707*	No
275,782	3,984,039	6,763,411	*11,023,232*	Pt 141	106,121	806,125	1,444,604	*2,356,850*	
913,174	*3,984,039*	*48,287,461*	*53,184,674*	Total	*431,207*	*806,125*	*14,803,225*	*16,040,557*	
Whole group				Pt 61	312,306	0	28,165,429	*28,477,735*	Yes
				Pt 141	169,661	3,177,914	5,318,807	*8,666,382*	
				Total	*481,967*	*3,177,914*	*3,3484,236*	*37,144,117*	

between an inherent data bias and an artifact induced by the necessary methodology of investigating what we are trying to investigate.

Inherent restriction to newer pilots in the CAIS data. Our CAIS data are restricted to newer pilots, since collection of school and examiner data only started for private pilots certificated after Jan. 1, 1995.

Table 3 shows that 26,696/63,951 of those pilots were IR (41.7%). This is a slightly lower percentage than the FAA's estimated private pilot IR percentage of 50.5%,[22] a circumstance for which we have no particular explanation.

Inherent restriction to newer pilots in the NTSB accident data. By design, we have imposed the exact same restriction on the NTSB accident data. Our original NTSB accident group started with the entire set of U.S. serious-to-fatal GA accident population during a specified time period. Entire populations have no selection bias *by definition*, because no one has been left out. Only subsets of populations can be biased, by excluding more individuals of one type than another.

The requirement that accident pilots also have School, Examiner, IR, and TFH data at least imposed the same constraint as that imposed on the CAIS data. As stated earlier, the vast majority of the 7,342-1,868 = 5,474 exclusions (74.6%) occurred because school and examiner data could not be retrieved from CAIS.

So, while we must logically restrict the conclusions of this study to newer pilots, this constitutes no particular fatal flaw to our study. We must simply not try to generalize the results of this study to pilots certificated before 1995.

Potential bias introduced during the PIC-classification process. We can also question whether there might have been any systematic bias in the process used to classify pilots-in-command. As stated, we used a "command status" hierarchy to assign PIC—a way of estimating who could/should take control of the aircraft, should something go wrong. Could that method of assigning PIC introduce biases that might also affect statistical analysis downstream?

To check, we can compare our PIC-selection methodology to a much simpler one, namely, the method of eliminating all accidents except for single-pilot flights. In a single-pilot flight, there is no dispute over who is PIC. Therefore, this is a plausible baseline against which to test our PIC selection method.

We first note that a large proportion of the accidents were known single-aircraft/single-pilot to begin with (1,775/1,868 = 95.0%). Hence, any alternate PIC-selection method will produce variation only for the remaining 5% of cases and is likely to be slight.

To test that, Table 5—now an analog of Table 1—shows frequency counts for single-pilot flights only.

We can compare Table 5's "actual" values, statistically, to "expected" values based on Table 1. To do this, we must first normalize Table 1 so that the cell totals we want to compare are equal. This is done by multiplying Table 1's whole-group values by 1,775/1,868, non-IR values by 1,002/1,036, and IR values by 773/832 to equate Table 1's 2x3 cell totals with Table 5's. Table 6 shows the result.

We can now compare each "actual" 2x3 in Table 4 with its *n*-equated (normalized) "expected" analog in Table 6 to statistically measure how much our "command status" method differed from a "single-pilot-only" method of determining PIC.

Fisher's Exact Test[23] yields *p*-values of .956, .992, and .988, respectively, all non-significant (NS). This suggests that our method of determining PIC did not significantly change the overall School x Examiner tabulation ratios. We therefore proceed with our analysis.

[22]Derived from www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2007/media/07-air1.xls, and www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2007/media/07-air10.xls,averaged over years 1998-2007, defined as $\Sigma N_{IR\ pilots} / N_{total\ pilots}$.

[23]Fisher's Exact Test is a more-precise substitute for the standard chi-square test, particularly useful when expected cell counts are < 5 (a violation of the assumptions of X^2).

Table 5. “Actual” single-pilot accident data, grouped by school, examiner, and instrument rating.

Examiner				School	Examiner				Instrument-rated at time of accident?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
16	0	1,602	*1,618*	Part 61	6	0	945	*951*	No
5	69	83	*157*	Part 141	1	17	33	*51*	
21	*69*	*1,685*	*1,775*	Total	*7*	*17*	*978*	*1,002*	
Whole group				Part 61	10	0	657	*667*	Yes
				Part 141	4	52	50	*106*	
				Total	*14*	*52*	*707*	*773*	

Table 6. “Expected” data (Table 1, normalized to single-pilot totals).

Examiner				School	Examiner				Instrument-rated at time of accident?
ASI	BYSCH	DPE	Total		ASI	BYSCH	DPE	Total	
15.2	0	1,593.5	*1,608.7*	Part 61	5.8	0	942.0	*947.8*	No ($p_{\text{Table 5 v 6}} = .992$)
5.7	74.1	86.5	*166.3*	Part 141	1.0	18.4	34.8	*54.2*	
20.9	*74.1*	*1,680.0*	*1,775.0*	Total	*6.8*	*18.4*	*976.9*	*1002.0*	
Whole group ($p_{\text{Table 5 v 6}} = .956$)				Part 61	9.3	0	653.1	*662.4*	Yes ($p_{\text{Table 5 v 6}} = .988$)
				Part 141	4.6	54.8	51.1	*110.6*	
				Total	*13.9*	*54.8*	*704.2*	*773.0*	

RESULTS

The Analytical Goal

Figure 1 shows the basic analytical structure. We sought to examine three major factors in private pilot instruction that might be associated with having an accident:

1. *School type* (Part 61 vs. 141).
2. *Examiner type* (ASI vs. BYSCHOOL vs. DPE).
3. *Instrument rating* (Was IR obtained after the private pilot certificate, Yes/No)

   while controlling for
4. *Risk* (Some metric based on TFH)

Figure 2 now illustrates how the data of Tables 1, 3, and 4 fit into our primary analytical structure.

As we shall see, the statistical method necessary to analyze these data will be somewhat involved. The most basic statistic that comes to mind for categorical data is chi-square (X^2). Chi-square would compare the “actual” accident data to a baseline of “expected” non-accident data, to tell us whether at least one cell in the School x Examiner x Instrument Rating matrix differed from the expected pattern.

However, several serious statistical considerations prevent the use of X^2. First, Figure 2 (left) shows that five of our 24 data cells contain fewer than five pilots—a violation of the assumptions of X^2. Four of these cells are “structural zeros” in our data matrix, because no Part 61 schools have authority to test their own graduates.

Second, X^2 cannot handle our intended first-pass risk-exposure continuous covariate of Total Flight Hours (Figure 2, right). Finally, X^2 cannot compute interactions between major factors, which we would like to test.

To investigate the effects of all our factors-of-interest, we need a more sophisticated multivariate statistical method. Log-linear analysis (LLA) is such a technique.[24]

Log-Linear Analysis

The basic method. Log-linear analysis (aka multiway frequency analysis; Norušis, 2012; Tabachnick & Fidell, 2001)[25] can handle the setup of Figure 2. The basic logic is the same as X^2, but the method is far more comprehensive. It produces results logically similar to regular analysis of variance (ANOVA) but works with frequency count data rather than the continuous scores required by ANOVA. Log-linear analysis will not only allow comparison of the kinds of data frequency matrices we have but can also control for covariates such as risk exposure, can calculate interaction effects, and is unfazed by structural zeros.

[24]Some readers may wonder whether odds ratios or logistic regression would be useful here. The answer is that odds ratios do address School and Examiner effects but not covariates. Logistic regression can address all effects, was tried, but failed to produce a useful model. The huge *n* of the non-accident data overwhelmed the small *n* of the accident data, resulting in a prediction equation that trivially assigned all cases as non-accidents.

[25]The Poisson is the appropriate modeling distribution to use with these data and is used.

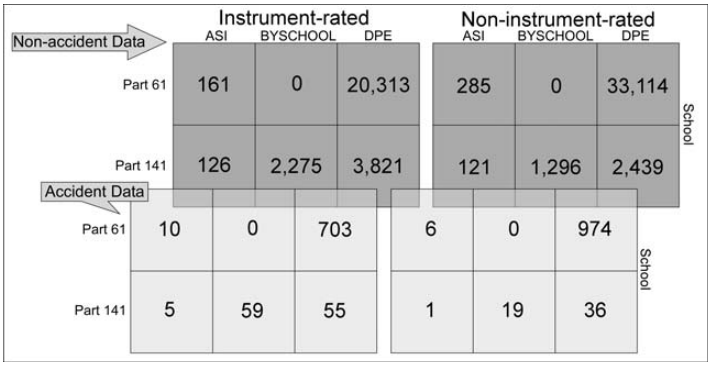


Figure 2 (left) now illustrates how the data of Tables 1, 3, and 4 fit into our primary analytical structure.

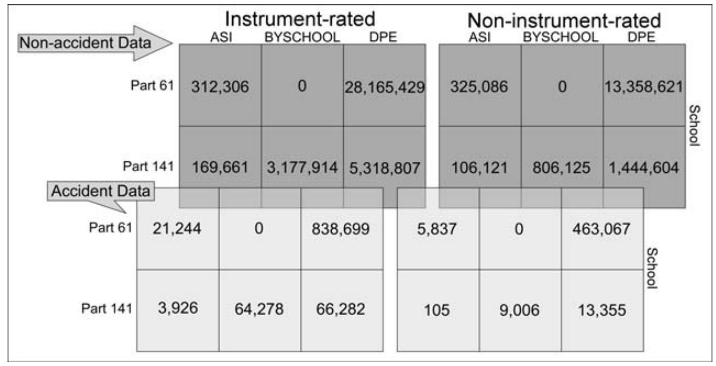


Figure 2 (right). The front 2x2x3 matrix represents aggregated accident data from Table 1. The rear matrix shows non-accident data from Table 3 (bottom). Corresponding TFH from Table 4, which will form the basis for a risk covariate.

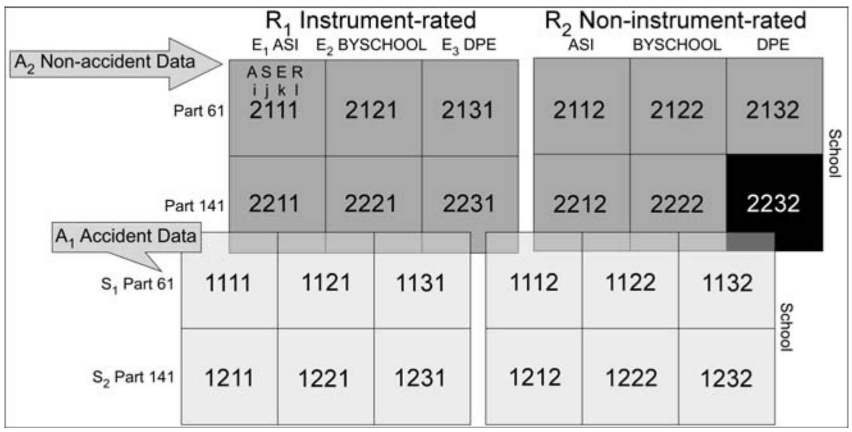


Figure 3. Cell subscripts associated with the 2x2x3x2 Accident x School x Examiner x Instrument Rating matrix representing Table 1 and Table 3. Cell_{2232} is the *reference cell* whose cell frequency count will be e^{u} (see text).

Given our data, LLA can partial out the effects of School, Examiner, Instrument Rating, and various interactions, while controlling for the effects of a risk covariate. It will do this by forming a set of 24 separate mathematical prediction equations, one per cell, to reconstruct the frequency counts of each cell in the overall 2x2x3x2 Accident x School x Examiner x Instrument Rating, *i,j,k,l* matrix implied by Figure 2.[26] Figure 3 shows the subscripts *i,j,k,l* associated with each cell.

Importantly, LLA will test its equations' *parameters*[27] for statistical significance, allowing us to estimate which variables and their interactions are reliably increasing or decreasing the individual cell frequency counts relative to one cell chosen as the *reference cell* (the black cell in Figure 3). Finally, through odds ratios (Hollander & Wolfe, 1999), LLA has the capability of telling us the *relative change in risk* posed by being a member of one group as opposed to another.

The main disadvantage of LLA is that results can be tricky to interpret. Multiple "significant" models are possible, given our data. So, the model we finally settle upon must be guided by a meaningful underlying logic. We do not simply run LLA with a saturated model (one including all main effects plus all possible interactions) the way we typically do with ANOVA. A saturated log-linear model will *always* fit the data perfectly, so we typically seek to eliminate as many statistically non-significant parameters as possible. This is an arcane point that we shall return to presently after some additional background information.

Understanding the mathematics. To completely understand LLA, we need to understand its mathematical logic, which differs from most other statistics. Equation 1 shows how each of our 24 prediction equations will be symbolized. By adjusting the shared parameters of these 24 equations, LLA's computational algorithm will try to make each cell's prediction equation duplicate that cell's observed frequency count.

For instance, for a model based on the Poisson distribution, containing all main factors plus all 2-way interactions, the *ijkl*th cell's predicted frequency count will equal (see Equation 1).

[26]The type of log-linear analysis used here estimates parameters by gradient-descent in multidimensional Poisson probability density function (pdf) space. Poisson distributions belong to a family of frequency distributions based on the natural logarithm *e* (Spanier, & Oldham, 1987). Such distributions share the useful characteristic that their indefinite integrals sum to 1.0, making them useful as pdfs. Specifically, a Poisson pdf is useful for predicting the likelihood of given values of discrete occurrences (.e.g., the probability of having 1, 2...*n* accidents), given a known, continuous maximum likelihood estimate (e.g., .001 accidents).

[27]A parameter is a weight or coefficient. Think of each parameter in Equation 1 as representing the influence of an independent variable (e.g., A) with unit value 1.0 multiplied times that cell's parameter for that variable (e.g., A_1).

$$\text{Predicted count}_{ijkl} = e^{\mu + A_i + S_j + E_k + I_l + R*R_{ijkl} + R_{Ai}*R_{ijkl} + A_iE_k + A_iI_l + A_iS_j + R_{Ek}*R_{ijkl} + R_{Il}*R_{ijkl} + R_{Sj}*R_{ijkl} + E_kI_l + S_jE_k + S_jI_l} \quad (1)$$

Here *e* is the natural logarithm (≈2.718), which is raised to a lengthy exponent. This exponent contains multiple terms (some of which may end up as zeros). Following basic algebra, any parameter in Equation 1's exponent whose value is greater than 0 will, therefore, *increase* that cell's predicted frequency count, while any parameter less than 0 will *decrease* it.

Describing first-order terms, the parameter μ (mu) represents a global constant added to every cell. The parameter A_i represents the "main effect of accident type," denoting either accident pilots (A_1) or non-accident pilots (A_2). S_j is the "main effect of school type," denoting either Part 61 (S_1) or Part 141 (S_2). E_k is the "main effect of examiner type," denoting either ASI (E_1), BYSCHOOL ("by school authority," E_2), or DPE (E_3). *I* is the "main effect of instrument rating," denoting either instrument-rated (I_1), or non-instrument-rated (I_2). *R* is a global coefficient representing the risk covariate, which will be individually multiplied by each *ijkl*th cell's unique aggregated risk covariate (the "sum of all risk" for every pilot in that cell, R_{ijkl}).

Describing second-order terms (2-way interactions), A_iS_j is a single number (*not* a multiplication of two separate numbers) representing the interaction of the *i*th *A* term (A_i) and the *j*th *S* term (S_j). A_iE_k performs a similar function for the interaction of *A* and *E*. A_iI_l, S_jE_k, S_jI_l, and E_kI_l behave similarly for their respective interactions.

As previously stated, every cell equation's exponent will contain the global constant μ. In practice, because terms in the exponent can assume values of zero, one cell's exponent will contain only μ, since all other terms zero out. That cell is designated as the *reference cell* (the black cell in Figure 3). Its referents are i_2 = non-accident, j_2 = Part 141, k_3 = DPE, l_2 =non-instrument-rated, $n_{ijkl} = n_{2232}$, and e^{μ} is defined as its cell frequency count.

Next, terms such as A_iS_j represent 2-way interaction effects unique to each cell.[28] Recall that an interaction term (e.g., A_1S_1) is not the multiplication of A_1*S_1. Rather, A_1S_1 is simply one number—the parameter representing an effect on all cells containing Part 61 pilots who also had an accident.

Finally, covariates are slightly more abstract. The term $R*R_{ijkl}$ represents a global covariate parameter, a single coefficient *R*, which will be multiplied by the cell in question's aggregated risk total (the "sum of all risk" for every pilot in that cell), to be described shortly. Additionally, a series of *R*-terms describe risk-interaction coefficients associated with A_i, S_j, E_k, or I_l. These are labeled R_{Ai}, R_{Sj}, R_{Ek}, R_{Il}, respectively, and are similarly multiplied by R_{ijkl} to help form the exponent of each cell's prediction equation.

So, this is the basic mechanism of general LLA. A set of parameters unique to a given model will be adjusted so that each of our 24 prediction equations will try to duplicate the actual pilot frequency count belonging to that cell. While the LLA procedure is running, a multidimensional parameter space will be generated, producing a same-dimensional error space, which can be globally minimized by gradient-descent numerical methods.

It is important to note that, unlike many other statistics: a) It is up to us to choose the *model*—the terms we want in Equation 1's exponent; b) Our statistical package (here, SPSS) will then follow a numerical error-minimization routine (SPSS, 1999) to arrive at values for parameters that best predict the real-world data, given our model; c) However, more than one "significant" model may exist. Therefore, the assumptions underlying each model's parameters are critically important.

Assumptions underlying our model. Log-linear parameters are abstract and require explanation. For instance, we designated the parameter *A* as the "main effect of accident type." In doing so, we theoretically assumed that there were a multitude of factors at work, which, taken together, represent *how common accidents are, relative to non-accident flights*. But, these factors are all lumped together, indistinguishably, so the term *A* tells us nothing specific about any single factor. One such factor might be "how well the pilot plans the flight." Another might be "how well she pays attention during landing." There could be hundreds of such influences on accidents. However, we are not interested in those particular details, so we combine them into the single parameter *A*.

What we need to clearly understand is that all *A* represents is *some adjustment to the exponent in Equation 1*. *A* is not a main effect in the sense we typically think of or care much about.

Likewise, the parameter *S* embodies "main effect of school type." However, like *A*, *S* is not something we care particularly about, all by itself. It merely embodies the *relative numbers of pilots* who graduate from Part 61 versus Part 141 schools. This, again, is merely a fact, just like the fact that accident flights are less common than non-accident flights. Table 3 showed us that far more pilots go to Part 61 schools. So, that *relative proportion* is what S represents—*not* whether one type of school shows a greater or lesser chance of graduates subsequently having an accident.

Similarly, the parameter *E* represents "main effect of examiner type." But, again, this is only a fact. Table 3 showed that far more pilots are tested by DPEs than ASIs, and that is all *E* stands for.

In LLA, the interaction parameter A_iS_j is actually the one that tells us whether getting one's private certificate from a Part 61 or Part 141 school is associated with subsequently more or fewer relative accidents. Likewise, the interaction parameter A_iE_k is the one that tells us whether pilots tested for their private certificate by an ASI or DPE are associated with subsequently more or fewer relative accidents.

So, unlike ANOVA, where main effects are of first-line importance, in our particular log-linear analysis, *interactions* are where we will first discover the kinds of effects we are most interested in.

Finally, as you might suspect, the 3-way interaction parameter $A_iS_jE_k$ might shed light on whether a particular *combination* of school and examiner has any effect. Unfortunately, there is a serious statistical issue with higher-order interactions. That issue is complicated, though, so we will postpone discussing it until Appendix B.

[28] Note that Equation 1 does not contain the 3-way interaction $A_iS_jE_k$. The reason why is explained in Appendix B.

The data. We use two types of data here. First, there are the frequency counts for accident and non-accident pilots, parsed by Accident, School, and Examiner, and set up as in Figure 2. The SPSS procedure we use (GENLOG) does not require normalization of the non-accident data. That is handled automatically by the GENLOG computational algorithm (SPSS, 2007).

Second, we have the covariate—our metric of risk exposure—sampled on Aug. 28, 2007.

Constructing the covariate. As one may imagine, detailed statistics on each pilot's risk exposure are not readily available. For one thing, risk is extremely complex and extremely hard to quantify. Additionally, many details about specific types of risk go unrecorded, since the task of keeping those kinds of detailed records would be quite costly. Finally, actual risk varies widely, depending on a host of factors such as the type of flight, phase of flight, and type of aircraft.

Although imperfect, total flight hours is a widely used proxy for risk exposure in aviation (Craig, 2001; Nakagawara, Montgomery, & Wood, 2002). Researchers usually assume that, in large samples, the statistical "noise" inherent in risk will average out, and that flight hours will correlate (covary) significantly with an underlying, theoretical construct of "true flight risk." This is a reasonable assumption. Nevertheless, we need to keep in mind that the correlation between flight hours and risk is far from perfect, so this measure of "risk" is a crude estimate at best.

When we speak of "TFH," it is important to distinguish between "total flight hours accrued over a pilot's career" versus "total flight hours accrued over some standard period of time," indicating a change ("delta", δ) in flight hours $\delta_{TFH} = TFH_{t2} - TFH_{t1}$.

A risk measure could arguably be better constructed from δ_{TFH} than from TFH accrued over a pilot's lifetime. Nonetheless, there are difficulties in trying to uniformly compute δ_{TFH} for all pilots, for instance: a) The length of such a "standard unit of time" is hard to establish; b) Many GA pilots do not fly regularly, and vary considerably in TFH, even over a fixed period of time; c) Some phases of flight (e.g., takeoffs and landings) are more dangerous than the cruise phase, yet most flight time is accrued in cruise, and; d) Formal, date-and-time-specific records of TFH are not uniformly and accurately kept by U.S. authorities. Instead, FAA and/or NTSB merely get a "snapshot" of TFH during medical certification, pilot certification, and/or at the time of an accident.[29] This snapshot is often not verified by checking against a pilot's logbook.

For U.S. non-accident pilots, the best available flight hour snapshots currently come from the FAA's Aerospace Medical Certification Division/Document Imaging Workflow System (DIWS), transcribed from Form 8500-8 gathered during pilot medical re-certification. For the private GA pilot, this involves Class-3 medical certification, recurring every 5 years for pilots under 40 years of age, and every 2 years thereafter.

Actual experience with these raw data reveals weaknesses relevant to our methodology. For one, it emerged that TFH reported during medicals are often estimates, rather than the exact current records taken from logbooks. This was confirmed by a Doctor of Medicine from AAM-630's Medical Research Team (Webster, 2010). In some cases, pilots even reported having fewer TFH at the time of their latest medical exam than at their previous medical, implying $\delta_{TFH} < 0$. Since δ_{TFH} would be the very metric used as a covariate, imprecision in either TFH_{t1} or TFH_{t2} contributes to imprecision in δ_{TFH}.

Moreover, using δ_{TFH} as a metric of risk assumes that risk is constant over the career of a pilot. To the contrary, we know it is not. Many factors affect the risk of a given flight. Notably, student pilots are typically at fairly low risk, because they have an instructor that is providing direct oversight. High-hour pilots are typically at low risk, because they are seasoned pilots. It is newly minted pilots who prove to be at greatest risk, statistically. Mathematically, the risk function is nonlinear, meaning risk is not a straight-line function of δ_{TFH}. It resembles more a skewed hump, tapering off at both extremes of TFH.

Therefore, to model this nonlinear risk, we decided to develop a new metric. This Advanced Risk Covariate (ARC) is detailed in Appendix A. At this point, ARC is *based* on TFH (not δ_{TFH}.), and simply calculates the actual chance of having an accident at a fixed, specific value of TFH, which we call a "point-estimate" of risk. Point-estimates are inherently less accurate than δ estimates but are easier to compute.

[29]Even though "past-90-day" flight hours are often kept, the 90-day estimate is too short a time to be statistically stable for our uses. A "past-365-day" estimate would be far more useful but is unavailable.

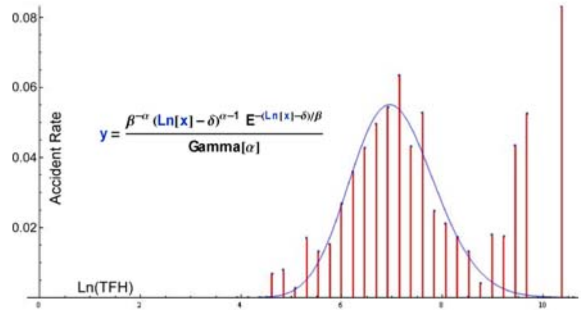
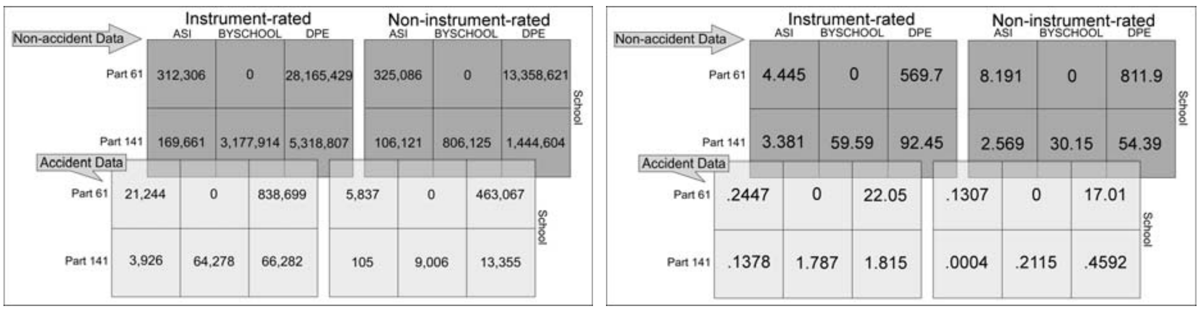


Figure 4. The Advanced Risk Covariate, a mathematical risk function based on our accident and non-accident data. It takes, as input, a value of TFH, and outputs an estimated accident rate, smoothing out noise in the data. There is a separate version for instrument-rated and non-instrument-rated GA pilots (IR data are shown).



a b

Figure 5. a) Aggregated TFH (from Figure 2, bottom); b) The corresponding Aggregated Advanced Risk Covariate. The AARC represents R_{ijkl} the "sum of estimated flight risk" for each cell.

Since we have both accident data and non-accident data spanning the same time period, ARC is derived from our actual data. Figure 4 illustrates a logarithmic plot of the basic function overlaid on one of our actual instrument-rated data groups.

Since LLA operates on aggregated data, individual pilots' values of ARC were summed to form an Aggregated ARC (AARC), as shown in Figure 5. The AARC then becomes the risk covariate used in LLA.

Summary of the final model. For the interested reader, Appendix B details the evolution of the log-linear modeling, with goodness-of-fit and parameter estimates, and walks through the logic of how the final model came to be. There, we also detail why we can effectively ignore the "main effects" of Accident (A), School Type (S), and Examiner Type (E), as well as the interactions not involving Accident.

For the sake of brevity, the final model is presented now, as Figure 6. To reiterate, the 2-way interactions involving Accident are where we will locate the effects of School, Examiner, Instrument Rating, and Risk, if any are significant.

This model consists of main effects plus all 2-way interactions of main effects *except* School x Examiner (S_jE_k) and Accident x School (A_iS_j), which were found to be insignificant. Mathematically, the general cell frequency count equation for this model is shown in Equation 2.

Here, the 2-way interactions of *AxS AxE* are the primary factors of interest, which we are directly tasked to investigate. *AxS* is absent from this model, because it was found in an earlier model to be insignificant (detailed in Appendix B). That meant that school-of-first-certificate had no significant effect on subsequent frequency of accidents, given these data.

$$\text{Predicted count}_{ijkl} = e^{\mu + A_i + S_j + E_k + I_l + R*R_{ijkl} + R_{A_i}*R_{ijkl} + A_iE_k + A_iI_l + R_{E_k}*R_{ijkl} + R_{I_l}*R_{ijkl} + R_{S_j}*R_{ijkl} + E_kI_l + S_jI_l} \quad (2)$$

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.
Constant	5.946	.583	10.202	.000
[Accident = 1]	-2.429	.625	-3.889	.000
[Accident = 2]	0[a]	.	.	.
[School_Type = 1]	1.682	.259	6.486	.000
[School_Type = 2]	0[a]	.	.	.
[Examiner_Type = 1]	-.885	.597	-1.483	.138
[Examiner_Type = 2]	.302	.410	.735	.462
[Examiner_Type = 3]	0[a]	.	.	.
[Instrument_rated = 1]	-1.307	.511	-2.557	.011
[Instrument_rated = 2]	0[a]	.	.	.
Adj_adv_risk_covar	.034	.011	3.186	.001
[Accident = 1] * Adj_adv_risk_covar	.095	.011	9.017	.000
[Accident = 2] * Adj_adv_risk_covar	0[a]	.	.	.
[Accident = 1] * [Examiner_Type = 1]	-2.543	.960	-2.647	.008
[Accident = 1] * [Examiner_Type = 2]	-.901	.650	-1.386	.166
[Accident = 1] * [Examiner_Type = 3]	0[a]	.	.	.
[Accident = 2] * [Examiner_Type = 1]	0[a]	.	.	.
[Accident = 2] * [Examiner_Type = 2]	0[a]	.	.	.
[Accident = 2] * [Examiner_Type = 3]	0[a]	.	.	.
[Accident = 1] * [Instrument_rated = 1]	1.557	.508	3.063	.002
[Accident = 1] * [Instrument_rated = 2]	0[a]	.	.	.
[Accident = 2] * [Instrument_rated = 1]	0[a]	.	.	.
[Accident = 2] * [Instrument_rated = 2]	0[a]	.	.	.

Parameter	Estimate	Std. Error	Z	Sig.
[Examiner_Type = 1] * Adj_adv_risk_covar	-.137	.051	-2.686	.007
[Examiner_Type = 2] * Adj_adv_risk_covar	-.004	.013	-.285	.775
[Examiner_Type = 3] * Adj_adv_risk_covar	0[a]	.	.	.
[Instrument_rated = 1] * Adj_adv_risk_covar	.005	.001	4.270	.000
[Instrument_rated = 2] * Adj_adv_risk_covar	0[a]	.	.	.
[School_Type = 1] * Adj_adv_risk_covar	-.031	.010	-3.069	.002
[School_Type = 2] * Adj_adv_risk_covar	0[a]	.	.	.
[Examiner_Type = 1] * [Instrument_rated = 1]	1.405	.485	2.900	.004
[Examiner_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.
[Examiner_Type = 2] * [Instrument_rated = 1]	.676	.354	1.910	.056
[Examiner_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 1]	0[a]	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 2]	0[a]	.	.	.
[School_Type = 1] * [Instrument_rated = 1]	-1.184	.198	-5.988	.000
[School_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.
[School_Type = 2] * [Instrument_rated = 1]	0[a]	.	.	.
[School_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.

a. This parameter is set to zero because it is redundant.
b. Model: Poisson
c. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

Figure 6. (top) Final-model parameters. This is detailed in Appendix B, including the goodness-of-fit test and residuals. Large colored areas contain significant values but do not address our primary focus of Accidents.

The interaction *AxE* represents the effect Examiner had on accidents. Our final model suggests that pilots tested for their first certificate by an ASI (cell A_1E_1) eventually had significantly fewer subsequent accidents than the reference group (DPEs, p=.008).

Importantly—*this particular finding is based upon an extremely small number of only 22 accidents* (see Figure 2, 10+5+6+1=22). As such, it is a textbook example of how "statistical significance" is not the same as "practical significance" to the conservative researcher mindful of the big picture. We have to ask ourselves that *if* we had access to all possible data, how likely would we be to get the same results? Given the extraordinary difficulty we encountered in matching pilots to data, spanning three separate databases that all had difficulty "talking to each other," *and*, given the extremely high data-loss rate (71.3%),[30] exactly how much practical significance should a prudent person assign to this one particular result? The circumspect answer is *"Little."*

Second, that the Aggregated Advanced Risk Covariate seems to relate significantly to accidents ("p=.000" in SPSS does not mean "zero probability"; it means "p<.0005"). Higher AARCs are associated with higher accident frequencies, which is what we hope for and expect from a risk metric. That alone does not establish AARC as a valid risk metric. It simply makes it arguably worthy of future investigation.

Third, in the A_1I_1 interaction, we see that instrument-rated pilots appear to have higher accident rates (p=.002). This is somewhat vexing, given that Appendix B shows that the model containing main effects and all 2-way interactions showed only near-trend for this effect (p=.106). We might well ask ourselves whether this is anything we ought to call meaningful, given that, as we began eliminating nonsignificant interactions, freed up variance could then be "fought over" by other parameters. That phenomenon (of previously insignificant factors becoming "significant" during backwards parameter elimination) is obviously a characteristic of, not just LLA, but of statistical modeling in general. And, it is one that we need to be wary of, as experienced critics of statistical methods.

Finally, we should note that, while it is possible in LLA to compute odds ratios for significant effects, we elect here not to do so, based on the argument that our input data were simply too stressed and/or sparse to take the analysis to that level of precision.

[30]65,819 usable pilots remained from an initial group of 302,685 (21.7%).

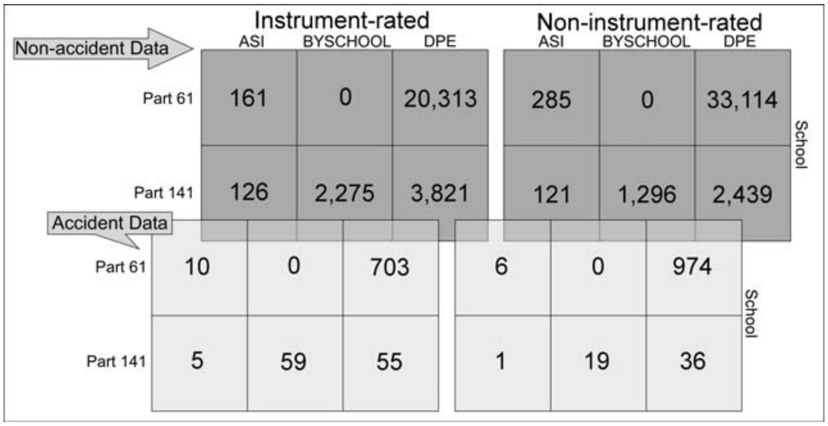


Figure 2. Data setup. The front 2x2x3 matrix represents aggregated NTSB accident data from Table 1, and is partially transparent, to show the rear matrix of aggregated FAA non-accident data from Table 3.

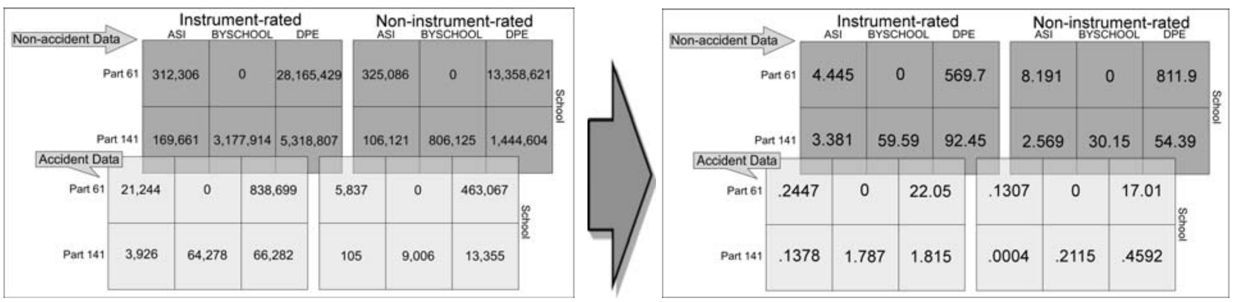


Figure 5. Total flight hours (at left) were used to produce a covariate representing known accident risk at various values of TFH. Values for individual pilots were then aggregated (at right) to form a total flight-risk value for each data cell.

DISCUSSION

Brief Summary of the Research Hypothesis, Methodology, and Results

This study was originally tasked to address what effects a general aviation (GA) pilot's type of education and certification testing might have on his or her subsequent flight safety record.

Given that there are many kinds of pilot instruction that could be tested, "education type" was operationalized as private pilot instruction in either a

- Part 61 or
- Part 141 school

"Certifying examiner type" was operationalized as pilots tested for their private pilot instruction by

- Aviation Safety Inspector (ASI),
- School Authority (BYSCHOOL, Part 141 graduates only), or
- Designated Pilot Examiner (DPE)

Because of the unavailability of earlier reliable FAA school and examiner records, results herein are restricted to pilots receiving their private pilot certificate after Jan. 1, 1995. No attempt should be made to generalize results to pilots certificated before then.

The experimental design implied by these factors-of-interest led to the following data setup, shown previously as Figure 2 and shown here again for convenience. Statistically, we compared frequency counts for NTSB accident data to a baseline of FAA non-accident data.

This experimental design compared 1,838 U.S. general aviation pilots involved in serious-to-fatal accidents during the time period 1/1/2003 to 8/26/2007 to a matched group of 63,951 non-accident U.S. GA pilots retrieved on Dec. 8, 2007.

To statistically help control for effects of pilot flight experience and flight-risk exposure on accidents, a) "Pilot experience" was operationalized partly as whether or not a pilot was instrument rated, and; b) Pilot total flight hours (TFH) were used to create a statistical risk covariate capable of predicting accident frequency based on TFH (see Appendix A). The Figure shown previously as Figure 5 illustrates.

Subsequent log-linear analysis produced the following main results:

1. Pilots who received their private pilot certificate from Part 61 schools were no more or less likely to subsequently have an accident than graduates of Part 141 schools ($p > .70$, NS).
2. Pilots who were examined by an Aviation Safety Inspector for their private certificate appeared less likely to subsequently have an accident than those examined by a Designated Pilot Examiner ($p < .01$). *However*, this result is suspect, because it was based on a total of only

22 accident pilots *and* because of the extremely high data loss rate (71.3%) prior to statistical analysis, which may have produced different results if more pilots could have been successfully matched to their school, examiner, instrument rating, and flight hours data.

Practical Significance of Results

The basic question of interest here was "Do first *training school type* and *certifying examiner type* affect a U.S. general aviation pilot's subsequent aviation safety record?"

The results of this study essentially imply that they do not. To the contrary—at least for GA pilots receiving the private pilot certificate from 2003-2007 and for whom data could be obtained—Part 61 graduates' subsequent accident rate appeared on a par with Part 141 graduates, and pilots tested by DPEs appeared equivalent to those tested under school authority. Graduates tested by ASIs showed a statistically lower accident rate, but this was based on a sample of only 22 pilots, rendering that result unreliable from a practical point of view.

RECOMMENDATIONS

Difficulties encountered during this project are elaborated in Appendix C. To summarize, the single greatest difficulty in trying to perform this study was trying to match pilots across the FAA and NTSB databases. The bulk of the problem stems from the lack of a common pilot reference designator (identification number) between NTSB and FAA records.

Past attempts have been made to integrate these kinds of data, the latest being the Bioinformatics Research Team of the FAA's Civil Aerospace Medical Institute development of a prototype "data warehouse." This was intended to assist in research efforts associated with statistical and epidemiological studies of the U.S. civil pilot population (Peterman, Rogers, Véronneau, & Whinnery, 2008). It incorporated NTSB and FAA Accident/Incident data, CAIS data, and medical certification data.

However, like many research efforts, this one may have been overlooked. So, if a recommendation were to be made on the basis of our experience with the present study, it would be the modest proposal that NTSB and FAA share what FAA calls their "UniqueID" designator, which allows the FAA CAIS and DIWS databases to "talk to one another." If this UniqueID could be extended to NTSB records, many problems that now exist trying to communicate between databases would disappear.

A second recommendation would be for the FAA to develop publically available user's manuals for CAIS and DIWS. NTSB currently has not only what amounts to such a user manual (their "data dictionary"), but also a completely searchable, pilot-deidentified accident database (www.ntsb.gov/avdata/) that can be downloaded and queried by anyone.

A third recommendation would be for the FAA to augment the flight hours information collected from pilots during their medical certification (FAA Form 8500-8). Particularly useful would be 12-month total flight hours, because this could form a fairly reliable and useful input to the Advanced Risk Covariate that was developed for this project (Equations 3, 4 in Appendix A).

REFERENCES

Air Safety Foundation. (2008). *2008 Nall Report*. Frederick, MD: Airline Owners and Pilots Association.

Broach, D. (2004). *Methodological issues in the study of airplane accident rates by pilot age: Effects of accident and pilot inclusion criteria and analytic strategy.* (Technical Report DOT/FAA/AM-04/8). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Craig, P. (2001). *The Killing Zone: How and why pilots die.* New York: McGraw-Hill.

Federal Aviation Administration (2010). U.S. Civil Airmen Statistics, Tables 1, 4, 10, 16. Downloaded Feb 2, 2011 from www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/2007.

Hollander, M., & Wolfe, D.A. (1999). *Nonparametric statistical methods* (2nd Ed.). (pp.477-483). New York: Wiley.

Knecht, W.R., Ball, J., & Lenz, M. (2010). *Effects of video weather training products, Web-based preflight weather briefing, and local versus non-local pilots on general aviation pilot weather knowledge and flight behavior, Phase 1*. (Technical Report DOT/FAA/AM-10/1). Washington, DC: Federal Aviation Administration, Research and Development Service.

Knecht, W.R., & Smith, J. (In review). *Effects of training school type and examiner type on general aviation flight safety*. (Technical Report). Washington, DC: Federal Aviation Administration, Research and Development Service.

Li, G. (1994). Pilot-related factors in aircraft crashes: A review of epidemiologic studies. *Aviation, Space, and Environmental Medicine, 65*, 944-52.

Nakagawara, V.B., Montgomery, R.W., & Wood, K.J. (2002). *The aviation accident experience of civilian airmen with refractive surgery*. (Technical Report DOT/FAA/AM-02/10) Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

National Transportation Safety Board. (2010). User-downloadable database. Downloaded May 26, 2011 from www.ntsb.gov/avdata.

National Transportation Safety Board. (2005). *Risk factors associated with weather-related general aviation accidents.* (Safety Study NTSB/SS-05-01). Washington, DC: NTSB.

Norušis, M. (2012). *IBM® SPSS Statistics 19 statistical procedures companion*. ch22 General log-linear analysis, 509-60. Upper Saddle River, NJ: Prentice-Hall.

Personal communication, L. Groff, Ph.D., Transportation Research Analyst, Office of Research and Engineering, National Transportation Safety Board, 13 Jan., 2011.

Peterman, C.L., Rogers, P.B., Véronneau, S.J.H., & Whinnery, J.E. (2008). *Development of an Aeromedical Scientific Information System for aviation safety.* (Technical Report DOT/FAA/AM-08/1) Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Spanier, J., & Oldham, K.B. (1987). *An atlas of functions*. New York. Taylor & Francis.

SPSS, Inc. (1999). *SPSS advanced models 10.0.* Chicago, IL: Author.

SPSS, Inc. (2007). *SPSS statistics 17.0 algorithms.* Downloaded Jun 6, 2011 from support.spss.com/ProductsExt/SPSS/ESD/17/Download/User Manuals/English/SPSS Statistics 17.0 Algorithms.pdf

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics (4th Ed.)*. Needham Heights, MA: Allyn & Bacon.

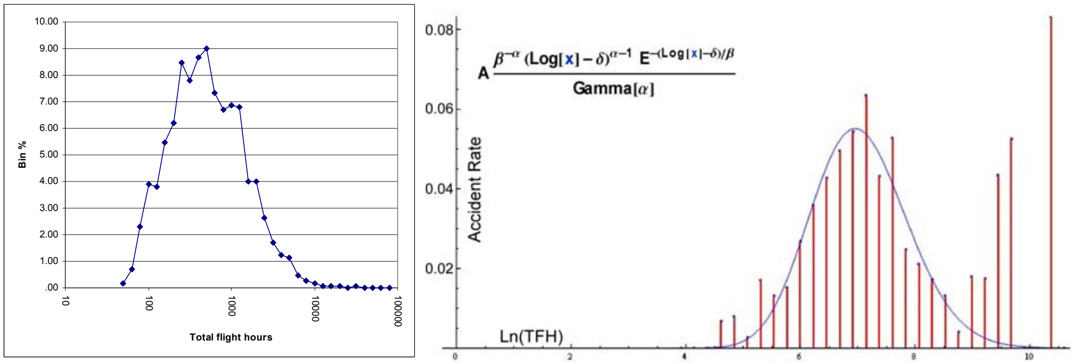
Webster, N. (Nov. 24, 2010). Private conversation, FAA AAM-630 Medical Research Team, Civil Aerospace Medical Institute, Oklahoma City,

Wolfram Research, Inc. (2008). *Mathematica, 7.0*. Champaign, IL: Wolfram.

APPENDIX A

Development of the Advanced Risk Covariate (ARC)

When raw pilot Total Flight Hours (TFH) was first tried as a risk covariate during preliminary log-linear analysis (LLA), it was found insignificant (p=.075). This prompted us to graph out our data's frequency distribution for TFH, shown in Figure 7a.



a b

Figure 7 (a). Frequency distribution of accidents for all pilots (combined instrument-rated and non-instrument-rated). The *x*-axis is the base-10 logarithm of TFH. The *y*-axis is the percentage of total accidents belonging to each *x*-axis frequency bin. (b) Similar data for instrument-rated pilots, graphed as accident *rates* (each bin's height equals the proportion of #accidents/(#accidents + #non-accidents).

Figure 7a is reminiscent of Craig's (2001) well-known book *The Killing Zone*, and shows, as he did, that the majority of GA accidents happen to pilots having *intermediate values* of TFH.

For our purposes, the basic problem is that LLA mathematically "wants" to interpret TFH values at one end or the other as "greater risk," which we can easily see is simply not true. It is the intermediate values of TFH that seem to have greater proportions of accidents. Figure 7b recasts our data as accident *rates*, which confirms this more firmly. Accident *rates* also form a humped distribution. This is particularly easy to visualize in the log(x) domain, which shrinks the long right-hand tail of the x-axis to more manageable dimensions.

The solution to this problem of "risk non-linearity" is to develop a metric that can justifiably be used as a true statistical covariate—one that directly expresses, as a scalar value, the *average probability of having an accident over a fixed period of time*, given one's TFH. Mathematically speaking, we want to express *risk = f(TFH)*, that is, "risk is a function of TFH." Then, we want to precisely define the function f to the highest degree of accuracy possible.

The difficulty in defining f, as Figure 7b shows, is noise in the data. Each data bin's y-value of risk is calculated easily enough, as $n_{accidents}/(n_{accidents}+n_{non\text{-}accidents})$ over some fixed period of time (as stated, ours was Jan 1, 2003 to Aug 28, 2007). However, some bins have very low numbers of pilots. This makes them much more susceptible to chance factors (a.k.a. statistical "noise"). The net result is difficulty fitting a mathematical modeling function to the data. We do not want the fitting process to be unduly influenced by extreme values belonging to small samples.

The solution lies in weighting the data. Before estimating f, what we can do is to generate a new data file having as many copies of each bin's accident rate as there were pilots that went into generating that rate. For example, if a particular bin's accident rate of 0.01 was generated by having 10 accidents and 990 non-accidents (10/(10+990) = 0.01), we then weight the data, effectively generating 1000 copies of the value 0.01 to represent that bin's contribution to the overall data. In this fashion, bins having extreme accident rates—but based on low numbers of pilots—will have less of an influence on the final estimate of f than they would if each bin's contribution were merely weighted equally as every other bin's.

The actual estimation of f is based on numerical methods. Numerical methods are mathematical techniques used to find solutions to complex problems in circumstances where no exact mathematical solution exists to a given problem. Ours is such a case. In our case, the procedure for finding f involves minimizing the total sums-of-squares [the sum of (each difference between the actual data and the prediction) squared] for the nonlinear gamma (Γ) probability density function. The method is complex, but is quite easily done by programs such as *Mathematica* (Wolfram Research, 2010).

The Γ pdf is chosen, based on work done by Knecht (in review), using NTSB accident data similar to those used here. Briefly, it was shown that a Γ pdf was capable of fitting eight sets of GA data taken from two time periods (1983-2000 and 2000-2007), two levels of accident seriousness (serious and fatal), and two levels of pilot instrument rating (instrument rated and non-instrument rated).

Using similar methodology, the present data were first parsed by instrument rating (IR versus non-IR). Accident rates were then calculated for the binned data and then presented to *Mathematica*'s *NonlinearModelFit* function to find solutions to the general form shown in Equation 3:

$$ARC_R = accident\ rate_R = A\frac{\beta^{-\alpha}\left(\ln(TFH)-\delta\right)^{\alpha-1}e^{-\left(\ln(TFH)-\delta\right)/\beta}}{\Gamma(\alpha)} \tag{3}$$

R Instrument rating (IR or non-IR)
A Amplitude
α Shape parameter of Γ_{pdf}
β Scale parameter of Γ_{pdf}
δ Location (shift) parameter
Γ (α)The value of the Euler gamma function at α

This resulted in the following parameter estimates:

$R_{IR}A \to 0.114905$ $\alpha \to 71.1448$ $\beta \to 0.0990758$ $\delta \to 1.3252*10^{-7}$
$R_{NIR}A \to 0.176691$ $\alpha \to 20.4944$ $\beta \to 0.34691$ $\delta \to 9.26536*10^{-7}$

Given these values, each pilot could now be assigned an ARC value, based on Equation 3. Since LLA is based on aggregated data, we then generated an Aggregated ARC (AARC) by simply summing the ARC values for all pilots within each of our 24 data cells (see Fig. 5). This AARC then substituted directly for what was originally TFH in LLA.

Limitations of the Method

The mathematically sophisticated reader will immediately spot the main limitation of this method, in that it is logically an "instantaneous approximation" (aka a point-estimate) of risk. As such, the ideal form for a better relative risk calculation would be:

$$ARC_{R\,\mathrm{int}} = \int_{TFH_{t1}}^{TFH_{t2}} ARC_R \tag{4}$$

where ARC_{Rint} is the definite integral of Equation 3, based on TFH from time t_1 to t_2, with t_1-t_2, of course, being equal for all pilots, and long enough to give a stable statistical estimate (say, 12 months).

The problem, naturally, is that we do not have $TFH_{t1\text{-}t2}$. In some cases, we can get TFH_{90days}, but that is not long enough to be a reliable indicator of true flight time. Therefore, what we are technically doing is basing AARC on the assumption that ARC is constant over a given time period and that aggregated data will be more stable than a single estimate. This is clearly not the best of all possible worlds, but, given our data, it is arguably the best we can do for the present.

Mathematica Code

Below is the *Mathematica* code used to weight the data and then parameterize Equation 3 for IR and non-IR pilots. Each data triplet in the "baseFile" (e.g., {717, 3.91, 0.0042}) represents the *i*th frequency bin's data, {n_i, ln(TFH_i), *accident rate*$_i$}.

```
baseFileNIR = {{717, 3.91, 0.0042}, {1377, 4.14, 0.0094}, {2421, 4.37, 0.0178}, {4736, 4.61, 0.0150}, {4297, 4.84, 0.0158}, {4545, 5.07, 0.0220},
    {3610, 5.30, 0.0274}, {5016, 5.53, 0.0249}, {2990, 5.76, 0.0344}, {2074, 5.99, 0.0497}, {2140, 6.22, 0.0379}, {1199, 6.45, 0.0484},
    {735, 6.68, 0.0653}, {766, 6.91, 0.0352}, {488, 7.14, 0.0676}, {405, 7.37, 0.0370}, {277, 7.60, 0.0469}, {386, 7.83, 0.0285}, {294, 8.06, 0.0204},
    {218, 8.29, 0.0183}, {230, 8.52, 0.0130}, {165, 8.75, 0.0364}, {87, 8.98, 0.0000}, {49, 9.21, 0.0408}, {27, 9.44, 0.0000}, {28, 9.67, 0.0000},
    {20, 9.90, 0.0500}, {18, 10.13, 0.0000}, {5, 10.36, 0.0000}, {3, 10.59, 0.0000}, {2, 10.82, 0.0000}, {0, 11.05, 0.0000}, {2, 11.28, 0.0000}};
baseFileIR = {{60, 3.91, 0.0000}, {115, 4.14, 0.0000}, {160, 4.37, 0.0000}, {293, 4.61, 0.0068}, {382, 4.84, 0.0079}, {705, 5.07, 0.0028},
    {1000, 5.30, 0.0170}, {2505, 5.53, 0.0132}, {2816, 5.76, 0.0153}, {2192, 5.99, 0.0269}, {2421, 6.22, 0.0359}, {1841, 6.45, 0.0429},
    {1549, 6.68, 0.0497}, {1860, 6.91, 0.0543}, {1479, 7.14, 0.0636}, {1389, 7.37, 0.0432}, {1172, 7.60, 0.0529}, {1534, 7.83, 0.0248},
    {1229, 8.06, 0.0212}, {1103, 8.29, 0.0172}, {1368, 8.52, 0.0132}, {743, 8.75, 0.0040}, {279, 8.98, 0.0179}, {57, 9.21, 0.0175},
    {23, 9.44, 0.0435}, {19, 9.67, 0.0526}, {14, 9.90, 0.0000}, {16, 10.13, 0.0000}, {12, 10.36, 0.0833}, {12, 10.59, 0.0000}, {7, 10.82, 0.0000},
    {4, 11.05, 0.0000}, {1, 11.28, 0.0000}};
workingFile = baseFileNIR;
If[workingFile == baseFileNIR, plotLabel = "baseFileNIR", plotLabel = "baseFileIR"];
LNAccRate = theWeights = theWeightsPlot = {};   maxLogFH = maxAccRate = 0;
(* Create a file of multiple data points, to weight each bin according to the n upon which the accident rate is based*)
For[i = 1, i ≤ Length[workingFile],  i++,
    If[workingFile[[i, 2]] > maxLogFH, maxLogFH = workingFile[[i, 2]]];
    If[workingFile[[i, 3]] > maxAccRate, maxAccRate = workingFile[[i, 3]]];
    AppendTo[theWeights, workingFile[[i, 1]]];
    AppendTo[LNAccRate, {workingFile[[i, 2]], workingFile[[i, 3]]}]
  ];
maxWeight = Max[theWeights];
For[i = 1, i ≤ Length[workingFile],  i++,
    AppendTo[theWeightsPlot, {workingFile[[i, 2]],  maxAccRate * workingFile[[i, 1]] / maxWeight }]
  ];
Clear[A, α, β, δ, constraints, initialValues];
gammaModel = A * PDF[GammaDistribution[α, β], x − δ];
selectedModel = gammaModel;
If[selectedModel == gammaModel,  modelName = "GammaModel"; constraints = {0 < A < 0.5, α > 0, β > 0, 0.2 > δ > 0};
     If[workingFile == baseFileNIR, initialValues = {{A, 0.17}, {α,  21}, {β, 0.34}, {δ, .00001}}];
     If[workingFile == baseFileIR, initialValues = {{A, 0.115}, {α,  71}, {β, 0.10}, {δ, .00001}}];
  ];
nlm = NonlinearModelFit[LNAccRate,   {selectedModel, constraints}, initialValues,   x,  Weights → theWeights,  AccuracyGoal → 4,  PrecisionGoal → 4,
    MaxIterations → 200 ,  Method → "Automatic" ]
(*, Gradient→"FiniteDifference"]  ]*)
If[ selectedModel == gammaModel ,     A = nlm[[1, 2, 1, 2]];    α = nlm[[1, 2, 2, 2]];    β = nlm[[1, 2, 3, 2]];   δ = nlm[[1, 2, 4, 2]];

    linearModel[x_] := A (β^-α (Log[x] − δ)^(α−1) E^(−(Log[x]−δ)/β)) / Gamma[α] ; (*This is  the linear version, using parameters est'd in the log domain *)

    Print["A→ ", A, "   α= ", α, "   β = ", β, "   δ = ", δ];(* n*binSize = the area under the curve *)
     ];
g1 = ListPlot[LNAccRate, PlotRange → All, Filling → Axis, Filling Style → Red];
g2 = Plot[nlm[x] , {x, δ, 11}, PlotRange → All, PlotStyle → {Thickness[.002], RGBColor[0, 0, 1]} ];
g3 = ListPlot[theWeightsPlot, PlotRange → All, Joined → True];
Show[g3, g1, g2, ImageSize → {800, 400}, AspectRatio → Full, AxesLabel → {"Ln[TFH]", "Accident Rate"}, AxesOrigin → {δ, 0},
   PlotLabel → plotLabel, GridLines → Automatic]
```

APPENDIX B

Evolution of the Log-Linear Modeling

The "Main-Effects-Only" Model

Equation 5 (below) represents the initial model with main effects only. Recall that we are trying to build a set of equations to reconstruct each *ijkl*th cell's frequency count. The subscript

- *i* represents Accident (A_i) (1=Yes, 2=No),
- *j* represents School (S_j) (1=Pt 61, 2=Examination by School Authority, 3=Pt 141),
- *k* represents Examiner (E_k)(1=ASI, 2=By School Authority, 3=DPE), and
- *l* represents Instrument Rating (I_l) (1=Instrument-rated, 2=Non-instrument-rated).

The global risk covariate is *R*, and R_{ijkl} is the Aggregated Advanced Risk Covariate (AARC) for cell_{ijkl}.

$$Predicted\ count_{ijkl} = e^{\mu + A_i + S_j + E_k + I_l + R^* R_{ijkl}} \tag{5}$$

Keep in mind that negative parameters (ones less than 0) have the effect of lowering cell count, while positive parameters increase the cell count.

Table 7a. The "main-effects-only" model, cell counts and residuals.

Cell Counts and Residuals[a,b]

Accident	School Certificate	Examiner Type	Instrument_rated	Observed Count	Observed %	Expected Count	Expected %	Residual	Standardized Residual	Adjusted Residual	Deviance
Yes	Part 61	ASI	Yes	10	.0%	18.074	.0%	-8.074	-1.899	-1.940	-2.076
			No	6	.0%	18.033	.0%	-12.033	-2.834	-2.886	-3.296
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	703	1.1%	603.776	.9%	99.224	4.038	5.052	3.934
			No	974	1.5%	597.301	.9%	376.699	15.413	21.460	14.112
	Part 141	ASI	Yes	5	.0%	6.026	.0%	-1.026	-.418	-.424	-.431
			No	1	.0%	6.013	.0%	-5.013	-2.044	-2.066	-2.537
		BySchExamAuth	Yes	59	.1%	115.666	.2%	-56.666	-5.269	-5.757	-5.822
			No	19	.0%	115.114	.2%	-96.114	-8.958	-9.529	-11.125
		DPE	Yes	55	.1%	194.426	.3%	-139.426	-9.999	-11.405	-11.830
			No	36	.1%	193.572	.3%	-157.572	-11.325	-12.370	-13.929
No	Part 61	ASI	Yes	161	.2%	249.847	.4%	-88.847	-5.621	-7.032	-6.016
			No	285	.4%	250.955	.4%	34.045	2.149	2.787	2.103
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	20313	30.9%	21381.116	32.5%	-1068.116	-7.305	-24.465	-7.367
			No	33114	50.3%	32446.898	49.3%	667.102	3.703	23.852	3.691
	Part 141	ASI	Yes	126	.2%	83.170	.1%	42.830	4.696	5.187	4.361
			No	121	.2%	82.882	.1%	38.118	4.187	4.531	3.915
		BySchExamAuth	Yes	2275	3.5%	1754.414	2.7%	520.586	12.429	18.380	11.880
			No	1296	2.0%	1663.806	2.5%	-367.806	-9.017	-13.817	-9.384
		DPE	Yes	3821	5.8%	3121.484	4.7%	699.516	12.520	19.265	12.092
			No	2439	3.7%	2916.427	4.4%	-477.427	-8.841	-15.723	-9.100

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar

Table 7b. The "main-effects-only" model, parameter estimates and goodness-of-fit.

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Constant	7.884	.017	465.401	.000	7.851	7.917
[Accident = 1]	-2.619	.034	-78.105	.000	-2.685	-2.553
[Accident = 2]	0[a]	.	.	.	.	.
[School_Type = 1]	1.098	.032	33.847	.000	1.035	1.162
[School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1]	-3.471	.046	-75.615	.000	-3.561	-3.381
[Examiner_Type = 2]	-.519	.021	-25.030	.000	-.560	-.479
[Examiner_Type = 3]	0[a]	.	.	.	.	.
[Instrument_rated = 1]	.002	.013	.162	.871	-.023	.027
[Instrument_rated = 2]	0[a]	.	.	.	.	.
Adj_adv_risk_covar	.002	.000	33.647	.000	.002	.002

a. This parameter is set to zero because it is redundant.
b. Model: Poisson
c. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar

Goodness-of-Fit Tests[a,b]

	Value	df	Sig.
Likelihood Ratio	1329.145	13	.000
Pearson Chi-Square	1219.703	13	.000

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar

This is not a fine-tuned model. The goodness-of-fit tests indicate that the cell-frequency predictions deviate significantly from the actual data, and high values for the residuals bear this out.

Consistent with our prior explanation of the logic underlying log-linear analysis (see Results/Log-linear analysis/Assumptions underlying our model), most main effects are significant ($p < .001$):

- There are significantly fewer accidents than non-accidents (A_1= -2.619, Z= -78.105, meaning "large decrease"). But, this is merely an expected fact.
- There are significantly more Part 61 graduates than Part 141 graduates (S_1=1.098, Z=33.847). This is also merely a fact.
- Many fewer pilots are tested by ASIs than by DPEs (E_1= -3.471, Z= -75.615), and fewer pilots are tested by school authority than by DPEs (E_2= -.519, Z= -25.03). Again, these are merely facts.
- The global risk covariate (AARC) is significant (R= 0.002, Z= 33.647). The Z>0 implies that higher values of global risk have higher cell frequencies. However, keep in mind that this effect applies to *both* accident and non-accident groups, so is not really a discriminator for accidents.

Instrument rating (I) is not significant in this model. However, keep in mind that we are only finding global influences on cell frequencies at this point—*not influences on accidents alone*. For instance, there is about the same ratio of non-IR to IR pilots in both accident and non-accident groups. From Tables 1 and 3, that odds ratio is (1036/832)/(37255/26696) = .892—not far from the statistically neutral ratio of 1.0.

The point is that "main effects" here are actually trivial and relatively uninteresting.

The "main-effects-plus-all-2-way-interactions" model

Logic and the previous model direct that we should include all main effects, because they will logically explain much of the variance, even though these are simply uninteresting facts. The next step is to add all 2-way interactions, represented by Equation 1 (which, as you will recall, earlier served as a prototype of the LLA methodology).

$$Predicted\ count_{ijkl} = e^{\mu + A_i + S_j + E_k + I_l + R^*R_{ijkl} + R_{Ai}{}^*R_{ijkl} + A_iE_k + A_iI_l + A_iS_j + R_{Ek}{}^*R_{ijkl} + R_{Il}{}^*R_{ijkl} + R_{Sj}{}^*R_{ijkl} + E_kI_l + S_jE_k + S_jI_l} \qquad (1)$$

Data-fit is perfect, as shown by the perfect frequency counts and zero-residuals in Table 8a.

Table 8a. The "main-effects-plus all-2-way-interactions" model, cell counts and residuals.

Cell Counts and Residuals[a,b]

Accident	School Certificate	Examiner Type	Instrument_rated	Observed Count	Observed %	Expected Count	Expected %	Residual	Standardized Residual	Adjusted Residual	Deviance
Yes	Part 61	ASI	Yes	10	.0%	10.000	.0%	.000	.000	.	.000
			No	6	.0%	6.000	.0%	.000	.000	.	.000
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	703	1.1%	703.000	1.1%	.000	.000	.000	.000
			No	974	1.5%	974.000	1.5%	.000	.000	.000	.000
	Part 141	ASI	Yes	5	.0%	5.000	.0%	.000	.000	.	.000
			No	1	.0%	1.000	.0%	.000	.000	.	.000
		BySchExamAuth	Yes	59	.1%	59.000	.1%	.000	.000	.000	.000
			No	19	.0%	19.000	.0%	.000	.000	.000	.000
		DPE	Yes	55	.1%	55.000	.1%	.000	.000	.	.000
			No	36	.1%	36.000	.1%	.000	.000	.	.000
No	Part 61	ASI	Yes	161	.2%	161.000	.2%	.000	.000	.000	.000
			No	285	.4%	285.000	.4%	.000	.000	.000	.000
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	20313	30.9%	20313.000	30.9%	.000	.000	.000	.000
			No	33114	50.3%	33114.000	50.3%	.000	.000	.000	.000
	Part 141	ASI	Yes	126	.2%	126.000	.2%	.000	.000	.	.000
			No	121	.2%	121.000	.2%	.000	.000	.	.000
		BySchExamAuth	Yes	2275	3.5%	2275.000	3.5%	.000	.000	.000	.000
			No	1296	2.0%	1296.000	2.0%	.000	.000	.	.000
		DPE	Yes	3821	5.8%	3821.000	5.8%	.000	.000	.	.000
			No	2439	3.7%	2439.000	3.7%	.000	.000	.	.000

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Accident * School_Type + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Examiner_Type + School_Type * Instrument_rated

As expected, most of the main effects are still significant. However, some of the explained variance now lies in the interactions.

Table 8b. The "main-effects-plus all-2-way-interactions" model, parameter estimates.

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Constant	5.984	1.037	5.771	.000	3.952	8.017
[Accident = 1]	-2.451	.967	-2.534	.011	-4.347	-.555
[Accident = 2]	0[a]	.	.	.	.	.
[School_Type = 1]	1.757	4.444	.395	.693	-6.953	10.467
[School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1]	-.982	1.622	-.605	.545	-4.162	2.198
[Examiner_Type = 2]	.318	.414	.768	.443	-.494	1.130
[Examiner_Type = 3]	0[a]	.	.	.	.	.
[Instrument_rated = 1]	-1.252	1.354	-.925	.355	-3.905	1.401
[Instrument_rated = 2]	0[a]	.	.	.	.	.
Adj_adv_risk_covar	.033	.019	1.752	.080	-.004	.071
[Accident = 1] * Adj_adv_risk_covar	.076	.292	.261	.794	-.496	.648
[Accident = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Accident = 1] * [Examiner_Type = 1]	-2.551	2.433	-1.048	.294	-7.320	2.218
[Accident = 1] * [Examiner_Type = 2]	-.929	.659	-1.409	.159	-2.221	.364
[Accident = 1] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 1] * [Instrument_rated = 1]	1.519	.939	1.619	.106	-.320	3.359
[Accident = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Accident = 1] * [School_Type = 1]	.241	.754	.320	.749	-1.237	1.719
[Accident = 1] * [School_Type = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [School_Type = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1] * Adj_adv_risk_covar	-.114	.274	-.415	.678	-.651	.423
[Examiner_Type = 2] * Adj_adv_risk_covar	-.005	.019	-.245	.807	-.042	.033
[Examiner_Type = 3] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Instrument_rated = 1] * Adj_adv_risk_covar	.005	.007	.651	.515	-.009	.019
[Instrument_rated = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[School_Type = 1] * Adj_adv_risk_covar	-.030	.014	-2.210	.027	-.057	-.003
[School_Type = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Examiner_Type = 1] * [Instrument_rated = 1]	1.342	1.514	.887	.375	-1.625	4.309
[Examiner_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 2] * [Instrument_rated = 1]	.693	.360	1.924	.054	-.013	1.398
[Examiner_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 1] * [Examiner_Type = 1]	-.202	5.922	-.034	.973	-11.808	11.405
[School_Type = 1] * [Examiner_Type = 2]	0[a]	.	.	.	.	.
[School_Type = 1] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[School_Type = 2] * [Examiner_Type = 1]	0[a]	.	.	.	.	.
[School_Type = 2] * [Examiner_Type = 2]	0[a]	.	.	.	.	.
[School_Type = 2] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[School_Type = 1] * [Instrument_rated = 1]	-1.096	1.210	-.906	.365	-3.467	1.276
[School_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.
b. Model: Poisson
c. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Accident * School_Type + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Examiner_Type + School_Type * Instrument_rated

This is a good opportunity to illustrate Equation 1. So, consider the modeling equation SPSS produces for $cell_{ijkl} = cell_{1111}$. Figure 5b supplies AARC=R_{1111}=.2447. Table 8b supplies cell exponents.

$$Predicted\ count_{ijkl} = e^{\mu+A_i+S_j+E_k+I_l+R*R_{ijkl}+R_{Ai}*R_{ijkl}+A_iE_k+A_iI_l+A_iS_j+R_{Ek}*R_{ijkl}+R_{Il}*R_{ijkl}+R_{Sj}*R_{ijkl}+E_kI_l+S_jE_k+S_jI_l}$$

$$Predicted\ count_{1111} = e^{\mu+A_1+S_1+E_1+I_1+R*R_{1111}+R_{A1}*R_{1111}+A_1E_1+A_1I_1+A_1S_1+R_{E1}*R_{1111}+H_{R1}*R_{1111}+R_{S1}*R_{1111}+E_1I_1+S_1E_1+S_1I_1}$$

$$= e^{5.984-2.451+1.757-.982-1.252+(.033*.2447)+(.076*.2447)-2.551+1.519+.241-(.114*.2447)+(.005*.2447)-(.030*.2447)+1.342-.202-1.096}$$

$$= e^{2.283} = 9.81$$

We can see in Table 8a that this produces (within rounding error) the observed value of $cell_{1111}$ = 10.

Accident-related effects. Interactions are where we expect to find any effects of School, Examiner, Instrument Rating, and/or AARC. Let us first examine the effects of the 2-way interactions on accidents.

Table 8b shows a non-significant Accident x AARC interaction (R_{Ai}=.076, p=.794), implying that our aggregated risk estimates are about the same for accident and non-accident pilots.

Likewise, all Examiner (A_iE_k= -2.551, -.929), Instrument Rating (A_iI_l=1.519), and School (A_iS_j=.241) interactions are non-significant. Recall these are the primary factors we originally set out to test.

Additional interaction effects. The only remaining effect that is statistically significant at the p<.05 level is the School x AARC interaction (R_{Sj}= -.030, p=.027). The negative value of R_{Sj} implies that pilots first-certificated from Part 61 schools may have pilots in a lower risk range than Part 141 schools. The reason is unclear.

Assessment of this model. The main peril of this "main-effects-plus-all-2-way-interactions" model is that it may be overfitted, meaning that it may have too many parameters, given the 24 data cells whose cell counts we are trying to fit. There are now six main effects (including the constant μ) plus 5(5-1)/2 = 10 interactions, making a total of 16 effects operating on 24 data cells.

Standard modeling procedure calls for removing some nonsignificant terms. One logical approach is to start eliminating non-significant 2-way interactions one by one, starting with the least-significant interaction, and monitor the effects on model fit and residuals. This is time-consuming but a conservative way to approach the situation.

Backwards Elimination of Nonsignificant 2-Way Interactions

Table 8b shows that the least-significant 2-way interaction is School x Examiner (S_jE_k, p=.973). Eliminating that still produces good model fit with low residuals, shown next in Table 9.

Table 9. Backwards elimination of nonsignificant 2-way interactions.

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Constant	5.955	.580	10.261	.000	4.817	7.092
[Accident = 1]	-2.426	.621	-3.909	.000	-3.642	-1.209
[Accident = 2]	0[a]	.	.	.	.	.
[School_Type = 1]	1.606	.321	5.010	.000	.978	2.234
[School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1]	-.931	.606	-1.535	.125	-2.119	.258
[Examiner_Type = 2]	.316	.411	.769	.442	-.489	1.122
[Examiner_Type = 3]	0[a]	.	.	.	.	.
[Instrument_rated = 1]	-1.295	.510	-2.539	.011	-2.295	-.295
[Instrument_rated = 2]	0[a]	.	.	.	.	.
Adj_adv_risk_covar	.034	.011	3.185	.001	.013	.055
[Accident = 1] * Adj_adv_risk_covar	.086	.026	3.325	.001	.035	.137
[Accident = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Accident = 1] * [Examiner_Type = 1]	-2.627	.980	-2.680	.007	-4.549	-.706
[Accident = 1] * [Examiner_Type = 2]	-.926	.652	-1.420	.156	-2.203	.352
[Accident = 1] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 1] * [Instrument_rated = 1]	1.546	.508	3.043	.002	.550	2.542
[Accident = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Accident = 1] * [School_Type = 1]	.224	.556	.403	.687	-.865	1.313
[Accident = 1] * [School_Type = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [School_Type = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1] * Adj_adv_risk_covar	-.123	.061	-2.009	.045	-.243	-.003
[Examiner_Type = 2] * Adj_adv_risk_covar	-.004	.013	-.331	.741	-.029	.021
[Examiner_Type = 3] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Instrument_rated = 1] * Adj_adv_risk_covar	.005	.001	4.170	.000	.003	.007
[Instrument_rated = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[School_Type = 1] * Adj_adv_risk_covar	-.030	.010	-3.053	.002	-.050	-.011
[School_Type = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Examiner_Type = 1] * [Instrument_rated = 1]	1.391	.484	2.872	.004	.442	2.341
[Examiner_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 2] * [Instrument_rated = 1]	.691	.356	1.940	.052	-.007	1.389
[Examiner_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 1] * [Instrument_rated = 1]	-1.136	.230	-4.940	.000	-1.587	-.685
[School_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.
b. Model: Poisson
c. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Accident * School_Type + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

Goodness-of-Fit Tests[a,b]

	Value	df	Sig.
Likelihood Ratio	.001	1	.973
Pearson Chi-Square	.001	1	.973

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Accident * School_Type + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

Cell Counts and Residuals[a,b]

Accident	School Certificate	Examiner Type	Instrument_rated	Observed Count	Observed %	Expected Count	Expected %	Residual	Standardized Residual	Adjusted Residual	Deviance
Yes	Part 61	ASI	Yes	10	.0%	9.976	.0%	.024	.007	.034	.007
			No	6	.0%	6.029	.0%	-.029	-.012	-.034	-.012
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	703	1.1%	703.024	1.1%	-.024	-.001	-.034	-.001
			No	974	1.5%	973.971	1.5%	.029	.001	.034	.001
	Part 141	ASI	Yes	5	.0%	5.023	.0%	-.023	-.010	-.034	-.010
			No	1	.0%	.972	.0%	.028	.029	.034	.029
		BySchExamAuth	Yes	59	.1%	59.000	.1%	.000	.000	.	.000
			No	19	.0%	19.000	.0%	.000	.000	.	.000
		DPE	Yes	55	.1%	54.977	.1%	.023	.003	.034	.003
			No	36	.1%	36.029	.1%	-.029	-.005	-.034	-.005
No	Part 61	ASI	Yes	161	.2%	161.001	.2%	-.001	.000	-.034	.000
			No	285	.4%	285.000	.4%	.000	.000	.034	.000
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	20313	30.9%	20312.999	30.9%	.001	.000	.034	.000
			No	33114	50.3%	33114.001	50.3%	-.001	.000	-.034	.000
	Part 141	ASI	Yes	126	.2%	126.000	.2%	.000	.000	.034	.000
			No	121	.2%	121.000	.2%	.000	.000	.034	.000
		BySchExamAuth	Yes	2275	3.5%	2275.000	3.5%	.000	.000	.000	.000
			No	1296	2.0%	1296.000	2.0%	.000	.000	.000	.000
		DPE	Yes	3821	5.8%	3821.000	5.8%	.000	.000	-.034	.000
			No	2439	3.7%	2439.000	3.7%	.000	.000	-.034	.000

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Accident * School_Type + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

From this, we see that Accident x School (A_iS_j, p=.687) is now the leading candidate for elimination (the Examiner x AARC interaction R_{Ek} is not eliminated, despite R_{E2}=.741, because R_{E1} is significant at .045).

Eliminating A_iS_j leads to the next (and final) model, shown below in Table 10 (and, previously, as Figure 6).

Table 10. Continuing backwards elimination to the next (and final) model (shown previously as Figure 6).

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Constant	5.946	.583	10.202	.000	4.804	7.089
[Accident = 1]	-2.429	.625	-3.889	.000	-3.653	-1.205
[Accident = 2]	0[a]	.	.	.	.	.
[School_Type = 1]	1.682	.259	6.486	.000	1.174	2.191
[School_Type = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1]	-.885	.597	-1.483	.138	-2.056	.285
[Examiner_Type = 2]	.302	.410	.735	.462	-.502	1.106
[Examiner_Type = 3]	0[a]	.	.	.	.	.
[Instrument_rated = 1]	-1.307	.511	-2.557	.011	-2.310	-.305
[Instrument_rated = 2]	0[a]	.	.	.	.	.
Adj_adv_risk_covar	.034	.011	3.186	.001	.013	.055
[Accident = 1] * Adj_adv_risk_covar	.095	.011	9.017	.000	.075	.116
[Accident = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Accident = 1] * [Examiner_Type = 1]	-2.543	.960	-2.647	.008	-4.425	-.660
[Accident = 1] * [Examiner_Type = 2]	-.901	.650	-1.386	.166	-2.176	.373
[Accident = 1] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Examiner_Type = 3]	0[a]	.	.	.	.	.
[Accident = 1] * [Instrument_rated = 1]	1.557	.508	3.063	.002	.560	2.553
[Accident = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Accident = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 1] * Adj_adv_risk_covar	-.137	.051	-2.686	.007	-.236	-.037
[Examiner_Type = 2] * Adj_adv_risk_covar	-.004	.013	-.285	.775	-.028	.021
[Examiner_Type = 3] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Instrument_rated = 1] * Adj_adv_risk_covar	.005	.001	4.270	.000	.003	.007
[Instrument_rated = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[School_Type = 1] * Adj_adv_risk_covar	-.031	.010	-3.069	.002	-.050	-.011
[School_Type = 2] * Adj_adv_risk_covar	0[a]	.	.	.	.	.
[Examiner_Type = 1] * [Instrument_rated = 1]	1.405	.485	2.900	.004	.455	2.355
[Examiner_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 2] * [Instrument_rated = 1]	.676	.354	1.910	.056	-.018	1.371
[Examiner_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[Examiner_Type = 3] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 1] * [Instrument_rated = 1]	-1.184	.198	-5.988	.000	-1.571	-.796
[School_Type = 1] * [Instrument_rated = 2]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 1]	0[a]	.	.	.	.	.
[School_Type = 2] * [Instrument_rated = 2]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.
b. Model: Poisson
c. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

Cell Counts and Residuals[a,b]

Accident	School Certificate	Examiner Type	Instrument_rated	Observed Count	Observed %	Expected Count	Expected %	Residual	Standardized Residual	Adjusted Residual	Deviance
Yes	Part 61	ASI	Yes	10	.0%	9.338	.0%	.662	.217	.381	.214
			No	6	.0%	5.851	.0%	.149	.062	.150	.061
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	703	1.1%	702.764	1.1%	.236	.009	.243	.009
			No	974	1.5%	974.298	1.5%	-.298	-.010	-.247	-.010
	Part 141	ASI	Yes	5	.0%	5.717	.0%	-.717	-.300	-.386	-.307
			No	1	.0%	1.093	.0%	-.093	-.089	-.101	-.091
		BySchExamAuth	Yes	59	.1%	59.000	.1%	.000	.000	.000	.000
			No	19	.0%	19.000	.0%	.000	.000	.000	.000
		DPE	Yes	55	.1%	55.180	.1%	-.180	-.024	-.207	-.024
			No	36	.1%	35.757	.1%	.243	.041	.220	.041
No	Part 61	ASI	Yes	161	.2%	161.900	.2%	-.900	-.071	-.404	-.071
			No	285	.4%	284.852	.4%	.148	.009	.404	.009
		BySchExamAuth	Yes	0	.0%	.000	.0%	.	.	.	.
			No	0	.0%	.000	.0%	.	.	.	.
		DPE	Yes	20313	30.9%	20312.998	30.9%	.002	.000	.074	.000
			No	33114	50.3%	33113.998	50.3%	.002	.000	.085	.000
	Part 141	ASI	Yes	126	.2%	125.045	.2%	.955	.085	.403	.085
			No	121	.2%	121.203	.2%	-.203	-.018	-.403	-.018
		BySchExamAuth	Yes	2275	3.5%	2275.000	3.5%	.000	.000	.000	.000
			No	1296	2.0%	1296.000	2.0%	.000	.000	.	.000
		DPE	Yes	3821	5.8%	3821.058	5.8%	-.058	-.001	-.405	-.001
			No	2439	3.7%	2438.946	3.7%	.054	.001	.403	.001

a. Model: Poisson
b. Design: Constant + Accident + School_Type + Examiner_Type + Instrument_rated + Adj_adv_risk_covar + Accident * Adj_adv_risk_covar + Accident * Examiner_Type + Accident * Instrument_rated + Examiner_Type * Adj_adv_risk_covar + Instrument_rated * Adj_adv_risk_covar + School_Type * Adj_adv_risk_covar + Examiner_Type * Instrument_rated + School_Type * Instrument_rated

which still produces good fit, and fairly low residuals.

Final Model

At this point, convention begs us to stop, the reason being that all main effects *must* remain included, plus, we see that all 2-way interactions now have at least one statistically significant component. Therefore, we choose to accept this as our final model, and are nearly ready to return to the Results section.

To summarize this final model, given these data:

- Most main effects were significant
    - There were fewer accidents than non-accidents
    - More pilots attended Part 61 schools than Part 141 schools for their first certificate
    - There were more non-instrument-rated (non-IR) pilots than instrument-rated (IR) pilots
- Some first-certificate-related factors appear associated with accident frequency
    - School-of-first-certificate was *not* associated with accident frequency
    - ASI as Examiner-for-first-certificate was associated with lower accident frequency
    - Having an instrument rating was associated with higher accident frequency
- The risk covariate (AARC) developed for this project significantly related to
    - Accident frequency(higher AARC was associated with greater accident frequency)
    - Examiner type(ASIs were associated with pilots with lower AARCs)
    - Instrument rating(IR pilots were associated with higher AARCs)
    - School type(Part 61 pilots were associated with lower AARCs)

Why We Avoid Analyzing Highest-Order Interactions

The primary problem with highest-order interactions (e.g., $A_iS_jE_kI_l$) in LLA, is that we can fit *any* data to a model containing *only* the highest-order interaction. As proof, we merely need consider that the highest-order interaction would actually be (in our case) a set of 24 individual coefficients, one per cell, free to vary for every individual cell. As such, it would be unaffected by any main effect or lower-order interaction. Ergo, the entire cell's frequency count could be duplicated by that one, unique parameter, rendering all others unnecessary, and trivializing any model based on it.

Caveats

Despite this array of seemingly careful methodology and advanced statistical techniques, one extremely important practical thing to keep in mind is the fact that, of our 24 data cells, four had 0 pilots, one cell had just 1 pilot, one cell had only 5 pilots, and another only 6. In practical terms, what that means is that—despite impressive-looking mathematics and 3-decimal-place significances—our results may be unstable, not because of our analytical method, but because of the quality of data input to that method. In other words, if we resampled the data, say from a slightly different time period, we might not get the exact same pattern of results.

This is a problem with the *data themselves*, not necessarily with the mathematics or SPSS. But, being careful and prudent researchers, it behooves us to honestly remind ourselves that instability may always lurk within small numbers whenever we sample those, no matter how careful we try to be or how meticulous our analysis.

That said, we can now return to the Results section.

APPENDIX C

A large number of issues emerged while using the NTSB and FAA databases. Some of these were minor and correctable, others were major and/or uncorrectable. However, all contributed significant difficulty to this project.

Issues With NTSB data

1. Missing data and/or data entry errors were common (sometimes blank cells, sometimes the numeral 0 where, for example, flight hours should be). We assume that all pilots involved in investigated accidents are in the database. But, data are often missing for a given pilot.
2. Pilots showing an accident event date earlier than their private pilot issuance date from CAIS. In these cases, it was unclear whether there was a data entry error, or perhaps the individual had had an accident while still a student.
3. Pilots listed as receiving their instrument rating on the same day as their private pilot certificate, or shortly thereafter which typically would not have been expected of pilots certificated during the data period analyzed (these turned out to be foreign pilots who were already instrument rated, who came to be U.S.-certified, and rapidly completed their examinations).
4. NTSB assigning one accident case number (the *ntsb_no* field) to each accident, no matter how many aircraft and/or pilots were aboard each aircraft. A naïve user may analyze data thinking that each row represents a separate accident.
5. Difficulty identifying the pilot in command (PIC). The NTSB has a field denoting pilot (*"PLT"*), as opposed to, for instance, co-pilot, student, or check pilot. It sometimes encodes *flight_type* as *PIC*, but this is typically the pilot at the controls. In most cases, that pilot truly is PIC—the pilot most responsible for managing the accident. However, in many cases, the researcher may not know of this field. In other cases, (e.g. student+instructor accidents or fatal accidents), the actual PIC may not be documented. NTSB staff are aware of this, and it is being discussed as an issue. However, documentation is not available to the public regarding this situation.
6. For what the NTSB reportedly claim to involve security reasons, at the time of this writing, FAA possesses only a circa-2007 copy of the NTSB accident database. Therefore, no research questions involving NTSB data beyond that time can be addressed without requesting a search by NTSB itself.
7. Difficulty identifying pilots' professions in an easily sortable way. How data have been entered into the database has reportedly changed several times
    a. Originally was "Yes/No," whether the pilot was a "professional pilot."
    b. This changed to "Y/N."
    c. Sometimes listed as "OP" ("occupation pilot," meaning "was a professional pilot") or "NOP" ("not occupation pilot").
    d. Sometimes listed as one of a limited number of options (e.g., *ct_crew_prof*, e.g., "aircraft mechanic," "clergy," "doctor/dentist," "farmer/rancher," "unknown").
    e. Field is often left blank, or listed as "N/A" (not applicable).

Issues With CAIS

1. No user's manual, list of frequently asked questions, or publically searchable database exists for large-scale research purposes. Whereas NTSB makes publically available both an explanation of what their data fields mean and a downloadable, queriable, "cleaned" version of their database (one that makes it extremely difficult to identify individual pilots), CAIS has no equivalent capabilities. Large-scale searches must be directed to FAA AFS-760 staff. And, while those staff were most helpful, several complications resulted:
    a. Not knowing what data fields were available ahead of time meant having to ask.
        i. In the case of the "UniqueID" (explained below), not knowing of its existence until late in the search process resulted in additional labor for both the authors and AFS-760 staff.
    b. Not knowing the limitations and/or unique characteristics of certain data fields led to some confusion and/or trial-and-error learning.
        i. Example 1: School and Examiner data only began to be collected starting in 1995.

ii. Example 2: Pilot certification numbers have changed over the years. For instance, prior to 2002, pilots' certificate numbers were their 9-digit Social Security Number. After 2002, for privacy reasons, the FAA started issuing 7-digit certificate numbers, and these were made available as an option for pilots already certified. Meanwhile, the certificate numbers listed with NTSB were not changed. Consequently, the same pilot's certificate number in CAIS often did not match their NTSB accident record certificate number.

iii. There is the possibility for mismatch between what the researcher imagines the data fields represent, versus what they may actually represent.

c. When problems arose with a given batch of data, AFS staff had to be consulted again.

2. Response to requests for CAIS data can take months, particularly if another organization has a large or high-priority project in progress. AFS-760 has limited staff, and projects must be priority-queued.

Issues With DIWS

1. DIWS also had no user's manual, list of frequently asked questions, or publically searchable database exists for large-scale research purposes.
2. Again, a member of AAM-300 must be contacted to perform the search query for the researcher.

Ultimately, a special identification number—the *UniqueID*—was mentioned by AFS-760 staff. This is an unadvertised ID number assigned to each pilot, which allows CAIS and DIWS to seamlessly match pilot records without having to resort to the potentially confusing pilot certification number.

Issues Common to Databases in General

1. Difficulty identifying a "GA flight"
   a. Different organizations define "GA" differently. For instance, one common FAA convention is "all N-tail-numbered aircraft not flying under Parts 121 or 135." However, inclusion of, say, aircraft above 12,500 lb may vary by research organization.
   b. These distinctions may be undocumented, or only locally documented, putting the uninitiated researcher at risk of obtaining search data inappropriate for their research question.
2. The AND-OR query problem
   a. If a database is searched on two or more fields with an AND query (e.g. return records containing A AND B), only records containing *both* fields will be returned. If there is a problem of any sort with either of the fields (e.g., missing data in one field), that record will not be returned.
   b. If a database is searched on two or more fields with an OR query (e.g. return records containing A OR B), records containing *either* field will be returned.
   c. The problem is that many researchers do not appreciate this distinction and the effect it has on the data that are subsequently retrieved.
3. Missing data
   a. Blank cells.
   b. Certain kinds of data not collected either before or after a certain date.
4. Data-entry errors (many kinds).