Measure for Measure

A statistician offers his perspective on the relative usefulness of different ways of measuring aviation safety.

BY ARNOLD BARNETT

There is no consensus about how best to measure the risk of flying. Recently, *The Wall Street Journal* used "fatal accidents per million departures" as its safety metric in a news story. Earlier *Journal* articles had cited statistics about "fatal accidents per 100,000 flight hours." The Boeing Co., not surprisingly, has long focused on "hull losses per million departures," although it has recently given equal emphasis to major events in which the hull was not destroyed.[1] The U.S. National Transportation Safety Board (NTSB) has calculated "passenger deaths per 100 million passenger miles," in part to facilitate comparisons with the safety of ground travel.[2]

This diversity among safety metrics raises several questions for the statistician. Given empirical evidence and common sense, which metrics are easiest to justify? Which are easiest to understand? As a practical matter, do all the metrics move up and down in unison? If so, trying to determine which one is the "best" might be a waste of time.

Given empirical evidence and common sense, which metrics are easiest to justify?

A quick visit to Google turns up nine primary safety metrics that have been used recently:

- Fatal accidents per 100,000 flight hours;
- Fatal accidents per million departures;
- Hull losses per million departures;
- Passenger deaths per 100 million passenger miles;
- Passenger deaths per million passengers carried;
- Passenger death risk per randomly chosen flight;
- Annual aviation death risk per million citizens;
- Accidents per 100,000 flight hours; and,
- Accidents per million departures.

Most of these statistics need no explanation, but some warrant further elaboration. "Passenger death risk per randomly chosen flight" is the answer to the question, "If a passenger chose a flight and seat at random from flights of interest — e.g., scheduled U.K. domestic jet flights in 1990–1999 — what is the probability he would not survive it?" "Annual aviation death risk per million citizens" is the ratio of a region's number of passengers killed in aviation accidents to its total population.[3] "Accidents" include all aviation events that cause death, serious injury or substantial damage. The great majority of accidents do not cause death.

Death is the most prominent common factor in the metrics above, appearing directly in seven of them. That emphasis seems sensible: if one assumes that *the air traveler's greatest fear is of being killed in a plane crash*, then statistics that reflect the likelihood of that outcome have intuitive appeal. Nonfatal injuries, terrifying near-accidents and massive property damage are certainly serious matters, but as a U.S. Supreme Court justice once said, death is different. Aviation metrics that suggest near-term mortality risk get closer to the issue of greatest interest than do other possible categories.[4]

The statistician recognizes that none of the indicators listed manifestly comes closest to the heart of the matter. To someone who believes that a 2,500-mi (4,023-km) flight from Sydney, Australia, to Perth entails far greater death risk than a 500-mi (805-km) flight from Sydney to Melbourne, a metric that treats flight length as irrelevant would seem deficient. To the person who believes that an upsurge in nonfatal accidents does *not* foreshadow a rise in fatal events, a safety indicator that is dominated by nonfatal accidents would seem lamentable.

Yet the statistician would also recognize that, unlike the choice of a favorite ice cream flavor, the selection of the best safety measure is more than a matter of personal taste. Every indicator listed above depends on one or more key assumptions. These assumptions can be tested against existing data, and when an axiom is inconsistent with the evidence, it undermines those metrics that depend on its accuracy.

Four General Truths About Aviation Safety

We will concentrate on passenger[5] deaths caused by aviation accidents, and will not consider terrorist and criminal acts. In evaluating specific risk indicators, four general points should be borne in mind.

1. *Passenger mortality risk on a flight is essentially independent of the flight's length or duration.*

The primary difference between long flights and short ones is that the former involve far more time at cruising altitude than the latter. But research at Boeing and elsewhere has demonstrated that only a small proportion of fatal air accidents are caused by crises at cruise altitudes. Other research has indicated that the average (intended) flight lengths for ill-fated airplanes are virtually the same as those for all airplanes.

Of the 15 scheduled U.S. domestic jet flights that resulted in fatal accidents from 1987 through 2006, only one was at cruise altitude when the emergency arose. Ninety-three percent occurred during the takeoff/climb or descent/landing phases of flight. Moreover, the flight distances of the segments that ended in fatal accidents were not especially large, averaging 626 mi (1,007 km), which is below the average segment length of approximately 750 mi (1,207 km) for all U.S. domestic jet flights over 1987–2006.[6]

What these patterns suggest is that all flight segments, regardless of length, entail nearly the same passenger death risk. Thus, an air journey from Montreal to Vancouver with intermediate stops at Toronto and Calgary is roughly three times as risky as a nonstop flight from Montreal to Vancouver. Yet the total distance traveled in the two itineraries is practically the same, as is the amount of time spent flying and the number of miles amassed by the traveler. This example suggests why using flight length, passenger miles or trip duration as the measure of passenger exposure to risk can lead to questionable inferences about safety.

2. *The category "fatal accidents" appears too broad for assessments about passenger mortality risk.*

The classification "fatal accident" makes no distinction between a crash that kills all 300 passengers aboard a plane and another event that kills one passenger out of 300. Thus, if a year with one accident that kills hundreds of travelers is followed by a year with two accidents that killed one passenger apiece, then risk would double under the criterion "number of fatal accidents."

Treating all fatal accidents alike would be appropriate if, once a life-threatening emergency has arisen, it is a matter of sheer luck how many perish. But a review of accidents suggests that it is not simply luck. Pilot skill can make a big difference. In one event in 1991, a Nigerian jet had to make an emergency landing at night. Because no available airport was near enough, the pilots had to put down in a field in the dark. Four passengers died in the crash, but 44 survived. At the former Eastern Airlines' terminal at JFK, a plaque memorialized the heroism of Capt. Charles White, whose plane suffered a midair collision over Connecticut. He managed a crash landing on a hillside. Three passengers out of the 45 aboard died, and the captain also perished as he tried to rescue a handicapped traveler from the burning wreckage.

Both of these events were fatal accidents. But is it irrelevant that 92 percent of the passengers (82 of 89) survived accidents that could well have killed everyone aboard? Moreover, the increased use of cabin floor lighting and fire-retardant materials aims to reduce fatalities in aircraft fires, even if it cannot eliminate them. Many observers believe, for example, that, but for improved precautions against fire, the death toll in the 1988 crash of Delta Air lines Flight 1141 would have been far greater than it was. However, because the event involved fatalities, the improved survival would not be reflected in fatal-accident statistics.

3. *The raw number of deaths in a fatal accident is an incomplete measure of the accident's safety implications.*

If an airliner hits a mountain, killing all passengers, the implications for system safety are not three times as large if 120 passengers are aboard rather then 40. And a crash that kills 15 passengers out of 15 does not have the same statistical meaning as one that kills 15 out of 250. In the latter case, excellent emergency procedures may have prevented a far worse outcome. Safety indicators that use raw numbers of deaths, in other words, are vulnerable to irrelevant fluctuations in the fraction of seats occupied, yet insensitive to salient information about the passenger survival rate.

Furthermore, one crash that kills everyone aboard a widebody jet might yield the same death toll as five crashes without survivors in smaller jets that are half full. One could argue that "a life is a life," and that the two scenarios involve the same degree of tragedy. It is not at all clear, however, that both scenarios say the same thing about the mortality risk of flying.

4. *The total number of major aviation accidents is a poor proxy for passenger mortality risk.*

It is sometimes suggested that the total number of accidents — fatal and otherwise — is a better barometer of system safety than statistics that focus on events that cause deaths. Because fatal crashes are mercifully rare, data about them can oscillate dramatically over time even in the absence of trends; the overall rate of accidents might be less susceptible to instability and thus might in principle be more informative.

One problem in using all accidents as a risk indicator is that, in some instances, a nonfatal accident might say more about the safety of the system than about the dangers it presents to passengers. In 1983, for example, an Air Canada Boeing 767 ran out of fuel at cruising altitude. The pilots made an emergency landing at an abandoned airstrip in Manitoba, damaging the airplane and causing some minor injuries, but avoiding any deaths. This event would be classified as an accident, as would a crash that killed everyone on board. But many people viewed what happened in Manitoba as more reassuring than horrifying.

Moreover, data analysis works against the notion that the overall accident rate is a "smoother" version of a risk statistic tied to deaths. Between the early 1970s and the mid-1980s, domestic U.S. jet accidents more than doubled while disastrous accidents — those that killed more than half the passengers on board — fell by a factor of *eight*.[7] Over 1990–1996 on major U.S. jet carriers, there was a *negative* correlation between an airline's rate of nonfatal accidents and the mortality rate among its passengers, i.e., airlines with more nonfatal accidents tended to have fewer deaths.[8] Every accident is of concern to aviation safety professionals, who must learn whatever they can from the event. But if the goal is to reflect the death risk that passengers face, then blurring the distinction between fatal and nonfatal accidents can be highly misleading.

Implications of the Four General Truths

How does it all add up? Every one of the nine risk metrics introduced earlier takes the form of a fraction, the numerator of which reflects the frequency and/or consequences of adverse events in aviation. In all but one of the fractions, the denominator is a measure of the amount of flying performed.[9] Thus, we effectively have a series of cost-benefit ratios, which differ, however, in how costs and benefits are measured.

The discussion above suggests that most of the numerators we have seen are flawed. Ratios that have number of accidents, number of deaths or total number of fatalities as their numerators *discard information about key events* that offers perspective about them. Most of the denominators seem flawed not because they use too little information, but because they use too much.

Of the nine risk measures, only one — passenger death risk per randomly chosen flight — avoids all the interpretive problems we have identified. It weights each crash by the *percentage* of passengers killed, meaning that a crash into a mountain killing all passengers is treated the same way whether the plane is half-full or completely full.[10] And the survival rate of a fatal accident fully enters the calculation. At the same time, risk exposure is measured on a per-departure basis, with no weight given to miles covered or hours in the air.

Transparency

Quite apart from their conceptual strengths and weaknesses, which of the indices just discussed are easiest to comprehend? We assume, as before, that the passenger is most interested in the risk that she will be killed on a forthcoming flight. How easy is it to infer a risk estimate from each index, even accepting it on its own terms?

Two metrics stand out as being intuitively accessible. "Passenger death risk per flight" and "passengers killed per million passengers carried" would seem the most transparent in estimating mortality risk, for each of them directly answers a question in the form of, "What are the odds?"

The other statistics appear less informative. The statistic "fatal accidents per million departures" falls short, for it says nothing about the chance of *surviving* an accident in which there are some fatalities. "Deaths per million flight hours" is incomplete because it does not indicate how many passengers landed safely over the million flight hours. The denominator of "deaths per million citizens" includes people who did not fly as well as those who did; hence, the metric says little about the risk to the air traveler. And the ratio "deaths per million passenger miles" would require adjustments in both numerator and denominator to generate a mortality risk statistic for, say, a 500-mi (805-km) flight.

Does It Matter?

The metric "passenger death risk per flight" (which is sometimes referred to as the Q-statistic) appears to get top marks in both conceptual soundness and transparency. Thus, if a statistician adheres to the four "general truths" above, he would likely conclude that the Q-statistic is the most attractive single metric of mortality risk. But we said earlier that if different safety indicators move the same way over time and across regions, then it doesn't matter much on which ones we focus. The statistician would therefore investigate with actual data whether the metrics move in parallel.

The prime statistical measure of whether two quantities move up and down together is the *coefficient of correlation*, which varies from minus 1 to 1. A coefficient near 1 means that the two quantities essentially move in lockstep: When one of them increases or decreases, it is all but certain that the other does the same. A coefficient near minus 1 implies opposite movements. When the coefficient is near zero, there is almost no relation between the movement of one quantity and that of the other. A coefficient around 0.5 typically means that the two quantities move the same way about 75 percent of the time and in opposite directions 25 percent of the time.

Table 1 concerns U.S. Federal Aviation Regulations Part 121 U.S. domestic flights — practically all passenger flights except air taxis — over the 20-year period 1987–2006. For every year, each of the nine safety metrics was calculated, and then the coefficient of correlation between each of the statistics and death risk per flight was computed. Each of the two metrics based on total accidents is *negatively* correlated with the Q-statistic, meaning that years in which accidents were relatively high tend to correspond to years in which mortality risk was low. The other metrics are positively correlated with death risk; because the coefficients fell in a narrow range around 0.5, however, the correlation is moderate but not strong.

In short, there is appreciable discrepancy between movements over time in death risk per flight and in the other metrics. At this point, there are two different ways one could proceed. One could argue that "death risk per flight" is the most defensible (or least objectionable) measure of passenger mortality risk, and adopt it as the primary statistic on the subject. Or —following the lead of the NTSB — one could release a "smorgasbord" of several of the listed statistics, and leave it to the reader to synthesize them to get an overview of passenger safety.

The statistician would be wary of the latter approach. When different statistics arise from contradictory starting premises, combining them to get a "holistic" impression has no clear logical underpinning. And it would be hard to justify any formal weighting scheme for the different statistics, as is suggested by the failure of attempts to create a "Dow Jones"–type index of aviation safety. Asserting that a synthesis of several

Mortality Risk Per Flight: Not Obvious From Most Statistics

Coefficient of Correlation Between Various Safety Metrics and Mortality Risk per Flight, U.S. FARs Part 121 Carriers, 1987–2006

Statistic	Correlation with Mortality Risk per Flight
Deaths per million passengers carried	0.56
Deaths per 100 million passenger miles	0.57
Deaths per million citizens	0.53
Hull losses per million departures	0.56
Fatal accidents per million departures	0.41
Fatal accidents per million flight hours	0.35
Accidents per million departures	– 0.18
Accidents per million flight hours	– 0.13

FARs = U.S. Federal Aviation Regulations

Source: Calculations by the author

Table 1

flawed statistics somehow transcends their deficiencies is a bit like saying that eight wrongs make a right.

Under these circumstances, we use the Q-statistic to assess patterns in passenger mortality risk.

Some Calculated Q-Statistics

Here we apply the Q-statistic in two ways: We consider scheduled commercial jet flights from 1960 onward, which is essentially the entire period during which passenger jet operations have taken place. We present the 1960–1999 data by decade, breaking the flights into four nonoverlapping categories, namely, developed world domestic; developing world international; between developed and developing world; and flights that begin and end in the developing world.

The calculated Q-statistics are shown in Table 2.

The key patterns in the data are obvious. Throughout the world and without *any* exceptions, jet travel has consistently become safer decade by decade. Overall jet passenger mortality risk fell by more than 90 percent between the 1960–1969 and 2000–2006 periods. The data offer no evidence that the percentage rate of improvement declined from decade to decade; this outcome is especially impressive because, as risk goes down, one might think that further improvement is harder to achieve. It is also apparent, however, that death risk is far lower on jet flights in the developed world than on those involving the developing world.

In assessing aviation safety metrics, the statistician would argue that no risk indicator should go unexamined, and that its underlying premises should be made explicit. When an indicator arises from premises that fare well under scrutiny, it is perhaps especially worthy of respect. The last thing the statistician would say is that, given that all safety indicators are imperfect, we are free to choose among them however we wish. If we lack an accurate understanding about present levels of safety, it seems less likely that we will be able to make flying even safer in the future. ●

Passenger Jet Travel: Safer by the Decade

Passenger Mortality Risk, Commercial Jet Aviation, 1960–2006

Period	Q-Statistic (Death Risk per Flight)
Developed World Domestic	
1960-69	1 in 1 million
1970-79	1 in 3 million
1980-89	1 in 4 million
1990-99	1 in 13 million
2000-06	1 in 70 million
Developed World International	
1960-69	1 in 400,000
1970-79	1 in 1 million
1980-89	1 in 4 million
1990-99	1 in 6 million
2000-06	1 in 9 million
Between Developed and Developing World	
1960-69	1 in 200,000
1970-79	1 in 300,000
1980-89	1 in 600,000
1990-99	1 in 1 million
2000-06	1 in 1.5 million
Within Developing World	
1960-69	1 in 100,000
1970-79	1 in 200,000
1980-89	1 in 400,000
1990-99	1 in 500,000
2000-06	1 in 2 million

Note: Statistics do not include crashes caused by criminal or terrorist acts. The calculations entail some approximations about the numbers of flights performed; see Barnett and Higgins[7] and Barnett and Wang[8] for discussions of the methodology and data sources used.

Source: Calculations by the author

Table 2

Arnold Barnett, Ph.D., is the George Eastman professor of management science at the Massachusetts Institute of Technology. He received the FSF President's Citation in 2002 for "harnessing the power of innovative statistical analysis to build conceptual bridges between data and significant aviation safety issues so that they can be readily understood."

Notes

1. The Boeing Co. *Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations, 1959–2006*. <www.boeing.com/news/techissues>.
2. Air Transport Association of America. Table, "U.S. Passenger Fatalities per 100 Million Passenger Miles," based on data from the NTSB. <www.airlines.org/economics/specialtopics/Airline+Safety.htm>.
3. U.S. Office of Hazardous Materials Safety. Table, "A Comparison of Risk: Accidental Deaths — United States — 1999–2003." <hazmat.dot.gov/riskmgmt/riskcompare.htm>.
4. Of course, an aircraft insurer might be most interested in hull loss rates. But we are taking the passenger's perspective in this article.
5. Mortality risks for on-board crewmembers are similar to those for passengers; however, data that would allow precise calculations for crewmembers are not available.
6. For example, the Economics Briefing of the International Air Transport Association, <http://www.iata.org/economics>, estimated the 2005 average sector length in U.S. domestic operations as 1400 km (about 870 mi). We use the more conservative figure of 750 mi because jet transport sector lengths have slowly increased in recent years.
7. Barnett, A.; Higgins, M.K. "Airline Safety: The Last Decade." *Management Science* Volume 35 (January1989), p. 7.
8. Barnett, A.; Wang, A. "Passenger Mortality-Risk Estimates Provide Perspective About Airline Safety." *Flight Safety Digest* Volume 19 (April 2000), Table 1, p. 3.
9. The exception is the denominator in "annual aviation death risk per million citizens."
10. See Barnett and Higgins (note 7) for the formula by which death risk per flight (the Q-statistic) is calculated.