AIRLINE SAFETY:
The Recent Record

NEXTOR Research Report RR-98-7

Arnold Barnett and Alexander Wang
Massachusetts Institute of Technology
Cambridge, Massachusetts

National Center of Excellence
in Aviation Operations Research

This report was prepared under the project "World Aviation Safety," which was supported by Federal Aviation Administration Contract DTFA03-97-D-00004 to the National Center of Excellence in Aviation Operations Research (Arnold Barnett, Principal Investigator). Statements in this report do not reflect the official positions or policies of the Federal Aviation Administration or the National Center of Excellence.

May 1998

AIRLINE SAFETY:
The Recent Record

ABSTRACT

Building on two previous MIT studies, we consider the mortality risk of passenger air travel over the decade 1987-96. We work with data from jet carriers around the world (except those in the former Soviet Union), and from commuter airlines in the US.

We present various relative and absolute analyses about airline safety, both over time and across carriers. Only rarely, however, do we discuss the circumstances of particular accidents: the aim is to measure the recent risk level rather than discuss how to reduce risk.

Our subjects of investigation include four groups of scheduled US flights: domestic jet services by long-established carriers, domestic jet flights by post-deregulation "new entrant" airlines, commuter air services, and international operations by US carriers. We examine too both domestic and international jet operations by First-World airlines outside the US, and corresponding operations in the Developing World. The specialized topics we consider include:

* how to measure passenger safety (and how not to do so)

* the volatility of mortality-risk data given the rarity of air crashes

* implications of the Valujet crash

* the comparative safety of auto trips and commuter flights

* the assertion that US airlines are "the safest in the world"

* an "ecological fallacy" that might cause exaggerated perceptions about the risk of flying Developing-World airlines

We note the overall safety of air travel: in the First World, a person who took one domestic jet flight every day would on average go 21,000 years before dying in a fatal crash. Even in the riskiest environments that we discuss, the daily air traveler would fly safely for more than a thousand years.

ARNOLD BARNETT and ALEXANDER WANG

Massachusetts Institute of Technology
National Center of Excellence in Aviation Operations Research

AIRLINE SAFETY: THE RECENT RECORD

I.INTRODUCTION

MIT researchers have performed two past studies about the mortality risk of air travel (Barnett, Abraham, and Schimmel [1979], Barnett and Higgins [1989]). The first effort covered the 20-year span 1957-76, and the second the decade 1977-86. Here we report on a third MIT analysis, which considers airline-passenger death risks around the world over 1987-96. This study attempts to provide a “baseline cardiogram” about worldwide aviation safety at a time when major initiatives are underway to improve that safety (e.g. the Global Airline Information Network).

Much as the second MIT study did more than update the original calculations, the present one goes substantially beyond replicating its predecessors. It considers some specific topics that, rightly or wrongly, did not seem of pressing importance when the previous studies were performed (e.g. commuter air safety). It explores at greater depth the fundamental issue of how best to measure the safety of air travel. And it offers some new methods for assessing how much meaning to attach to apparent trends in the data.

When the decade 1987-96 ended, aviation safety was a topic of immense concern in the US. In an end-of-1996 Associated Press survey, US newspaper editors and television news directors voted the July crash of TWA 800 as the “biggest” news story of 1996, and the May crash of Valujet 592 as the fifth biggest story. These events generated huge repercussions within and without the FAA that are still being felt today. These repercussions, in turn, produced fears that air safety planning in America has become dominated by “the latest headlines.” A major virtue of a study like this one is that, by its nature, it offers a broad perspective on the safety of air travel, calling attention to the accidents that didn’t occur as well as the tragedies that did.

While it is traditional in empirical reports to summarize key findings in the introduction, doing so risks oversimplification in the present context. But some general points we can make at the outset are:

* Over the most decade 1987-96, there was no net gain in worldwide aviation safety, and no net progress even among the airlines with the greatest room for improvement. This outcome contrasts sharply with that for the previous decade, during which passenger mortality risks fell by factors at least five compared with earlier years.

* Barnett and Higgins (1989) reported that established US domestic jet carriers could “continue to lay claim to being the safest group of airlines in the world.” That claim is harder to advance today because, in a statistically-significant performance surge over the previous decade, domestic jet carriers of other First-World nations achieved lesser mortality risk over 1987-96 than did their American counterparts.

* Barnett and Higgins also argued that, while US domestic jet travel had become safer after airline deregulation in 1978, the improvement might have been greater still in the absence of deregulation. This assessment was based on a significantly higher

passenger mortality risk over 1979-86 on “new entrant” jet carriers than on established airlines. Despite the Valujet crash, however, it seems less clear today that new entrant airlines can meaningfully be identified with lesser safety.

* Beyond such comparative issues, however, First-World domestic jet safety is exceedingly strong in absolute terms. Passenger death risk is about one in 8 million on First-World domestic jet flights; at that rate, a passenger who took one such flight every day could on average go for 21,000 years before succumbing to a fatal crash.

* Passenger death risk over 1987-96 was about 15 times as high on domestic jet flights in the Developing World as in the First World. However, the discrepancy is somewhat less on international jet routes and, on the subset of routes on which First and Third-World airlines directly compete (e.g. Paris-Karachi), the difference virtually disappears. Operations in the Developing World appear to entail higher risks for all airlines that perform them, and not just local flag carriers.

These and several other findings are subjects of detailed discussion in this report. All will require careful qualification. (We are not suggesting, for example, that domestic US jet carriers have demonstrably “fallen behind” those elsewhere in the First-World; what we will describe is the diminishing basis for claims of outright US superiority.) In addition to describing our analyses, we present much of the raw data used in our calculations so that readers are in a position to perform alternative ones.

We start our work in the next section, where we discuss several approaches to measuring air safety (Section II). Then we proceed to assess passenger death risk over 1987-96 for established US domestic trunklines (Section III), other US jet operations (Section IV), US commuter propjet services (Section V), First-World carriers outside the US (Section VI), and Developing World airlines (Section VII). We present some details about the calculations in two Appendices.

II. MEASURING AIRLINE SAFETY

Aviation safety is such a broad topic that we must limit the scope of this manuscript to keep it finite. Because the air traveler’s greatest fear is of being killed in a plane crash, we concentrate here on statistics about the likelihood of that outcome. We study absolute and comparative passenger death risks on different airlines over the last decade. (Airlines seem appropriate as units of study because each is a decision-making body with responsibility for the safety of its passengers.) Only rarely do we discuss the circumstances of particular accidents; the aim is more to assess the level of risk than the reasons the risk assumes that level.

Some Mortality-Risk Measures

Once attention is restricted to airline mortality risk, the question becomes how best to measure that hazard. That the issue is not straightforward is suggested by the fact that several proposed measures--some of them very recent--appear to suffer serious shortcomings. Consider, for example, a statement that appeared in *The Wall Street Journal* on 8/11/97:

"NTSB (National Transportation Safety Board) studies show that, from 1993 to 1996, scheduled US carriers averaged only 0.2 fatal accidents per 100,000 flying hours, less than half the fatal accident rate for the four-year period a decade earlier."

The statistic "fatal accidents per 100,000 flying hours" is problematic in both numerator and denominator. The phrase "fatal accident" seems to blur the distinction between an event that kills one passenger and a crash that kills everyone aboard, and would apparently not give credit to safety improvements (e.g. in fire-retardant materials) that reduce casualties but do not prevent them. And, given that the overwhelming majority of accidents occur during the takeoff/climb and descent/landing phases of flight (e.g. see Boeing (1996)), statistics tied to the duration of air trips are of uncertain value. If the mean trip time changes from one period to another, then the proposed safety indicator could change for reasons having nothing to do with safety.

While any indicator is imperfect, raw statistics suggest that the concerns above are more than mere quibbles. Over 1983-86, three US jets suffered crashes that killed the heavy majority of those on board. Over 1993-96--with a slightly higher number of flights--the corresponding statistic was five. Thus, the fact that the index"fatal accidents per flying hour" discerned a factor-of-two improvement in safety between 1983-86 and 1993-96 may say more about its own limitations than about the actual state of affairs.

Airline "Report Cards"

Another mortality-risk measure was suggested in 1997 by the Air Travelers Association, which issued safety "report cards" about airlines around the world (http://www.1800airsafe.com). For each carrier, the association calculated a numerical safety score S for 1987-96:

$$S = 100 - (10{,}000)F/N$$

where:

F= number of events with passenger fatalities over 1987-96
N= number of flights (in 1000's) over 1987-96.

Then the scores were converted into letter grades: 90-100 was A; 80-90:B ; 70-80:C; 60-70:D; and below 60: F. For example, an airline with two fatal events and 1.5 million flights would have an S-score of 100- [20,000/1500] = 86.7, and would receive a grade of B.

A positive aspect of the S-score is that it ignores flight duration and follows the more defensible "one flight, one vote" rule. But it follows NTSB in treating all fatal events as equivalent, which allows an airline that has two fatal events with very few deaths to get a lower grade than a slightly-larger airline that had two crashes with no survivors.

There is, however, a deeper deficiency in the report cards. Because accidents that cause deaths are mercifully rare, an airline's fatal-accident rate over a limited period is but a hazy indicator of its long-term rate. The report cards, however, not only fail to mention this point but are gravely compromised by it. A small airline that has 20,000 flights per year gets an A if it had no fatalities over 1987-96; if it suffered just one death on one flight, however, its S-score falls to 50 and its letter grade plunges to F. Because of peculiarities like this, statistically-meaningless differences between carriers can generate extreme disparities in the way the Air Passengers Association describes their safety records.

Another measure of death risk is the ratio of passengers killed to passengers carried ("deaths per enplanement"), which has the advantage that it can be interpreted as a probability. A difficulty arises, however, with its use of raw numbers of deaths as its numerator. As Barnett and Higgins noted:

"When a Boeing 727 hits a mountain killing all passengers, the implications about safety are not three times as grave if there were 150 on board rather than 50. And a crash that kills 18 passengers out of 18 should be distinguished from another that kills 18 out of 104. (In the latter case, the high survival rate might reflect excellence in the airline's emergency procedures.) Statistics that weight crashes solely by their numbers of deaths, in other words, are vulnerable to irrelevant fluctuations in the fraction of seats occupied yet insensitive to salient variations in the fraction of travelers saved."

<u>The Q-statistic</u>

The MIT research has used a different death-risk measure based on the question: if a passenger chooses a (nonstop) flight completely at random from among those of interest (e.g. US domestic jet services over 1990-95), what is the probability that he will be killed during the flight? (By "completely at random," we mean that, if there were N flights of interest, the chance is 1/N that the passenger would select any particular one.) Simple probabilistic arguments establish that Q follows:

$$Q = \underline{\quad} x_i / N \qquad (2)$$

where x_i = proportion of passengers on the i^{th} flight (i=1,2,..,N) who do not survive it. (If, for example, the flight lands safely, then x_i =0.)

Weighting accidents by the fraction of passengers killed (as Q does) seems more illuminating than using the number who perished (without reference to how many were on board) and greatly more illuminating than reducing flight outcomes to the yes/no question "did any passengers perish?" Furthermore, Q adapts the best features of the previous three indicators: it ignores the length and duration of individual flights, is easily calculated, and is easy to understand. For these reasons, the Q-statistic will be the "workhorse" of this analysis.

Whatever its virtues, however, the Q-statistic cannot circumvent a basic fact: given the infrequency of fatal air accidents, the data about them suffer the "loud" statistical noise that accompanies all rare phenomena. We will be discussing this issue in detail; right now, however, we turn to an approach to evaluating airline safety that is drawing increased attention.

A Broader Data Set

It has been suggested that safety assessments about a particular airline should be based on all untoward events it has experienced over a period, rather than the subset that led to passenger fatalities. The idea is that is largely a matter of luck whether such an event becomes a disaster or an incident; hence, even if the focus is on future mortality risk, it is wiser to use the overall accident/incident rate as a proxy than the actual recent death risk. Among other things, such a broadening of the data base increases "sample sizes" and thereby diminishes statistical volatility.

Interest in this viewpoint grew--indeed, surged--after the 1996 crash of Valujet Flight 592 that killed all 104 passengers. The ensuing debate about Valujet's safety seemed to revolve less around the crash of Flight 592 per se than the airline's long string of nonfatal accidents and incidents that preceded the disaster. More recently, the FAA has made available on the World Wide Web a full listing of mishaps that befall each US airline, and has been under some Congressional pressure to use the data to rank these carriers in safety. In May, 1996, the FAA released a report about domestic airline safety that compared various accident/ incident statistics for individual US jet carriers; the report made no reference, though, to which of the events studied had caused deaths.

The notion of emphasizing all mishaps and not just fatal crashes has some appeal; it falters badly, however, under scrutiny. There is the immediate issue of data availability and reliability. In international comparisons, it is hard enough to obtain data about fatal crashes; learning about those without deaths might be exceedingly difficult. Even in regions that have reasonably clear standards for reporting nonfatal mishaps, there is room for interpretation, and some carriers may report more precisely than others. It would be unfortunate if an extremely-conscientious airline that reported all its mishaps artificially seemed particularly prone to them.

Moreover, it may be a bit glib to say that it is a "matter of chance" whether a mishap turns disastrous, as is suggested by two examples. The first involved an Aloha Airlines Boeing 737 that suffered an in-flight structural failure that practically destroyed the upper half of its fuselage. A flight attendant died when she was blown out of the crippled plane, but the pilot managed to land the plane with no passenger deaths. The second took place on an Air Canada 767 that, because of a misunderstanding about whether its kerosene requirement was expressed in pounds or kilograms, literally ran out of fuel in mid-air. The pilots brought it down safely on an abandoned airstrip in Manitoba. While the plane was damaged in the highly-irregular landing, no passengers were seriously hurt.

Both these events meet a broad definition of accidents. But it was not luck but rather two outstanding cockpit crews that saved the passengers from airborne crises that could easily have killed them all. Indeed, one could argue that the consequences of an accident may say more about the safety of an airline's operation than does the existence of the accident. Yet such consequences get no weight in overall accident/incident statistics.

Moreover, data analysis fails to support the conjecture that, the greater an airline's involvement in mishaps, the greater its propensity to suffer the disasters that passengers fear. Between the early 1970's and the mid-1980's, reported accidents and incidents per 100,000 flights doubled on major US domestic airlines; passenger death risks, however, did not double, but rather fell by a factor of three (Barnett and Higgins, [1989]; hereafter BH). More recent data from the period 1/1/90-3/31/96--the very time span of the FAA's 1996 report--suggest the dangers of using carrier-specific mishap data as a proxy for mortality risk.

All US jet fatalities over that period occurred on carriers classified as "major" by the FAA. Table 1 below presents cross-airline correlation coefficients between various statistics about nonfatal incidents/accidents per 100,000 flights and passenger death risks as measured by Q-statistics. As we see, all the correlations are negative. And, as one tries to weed out the less serious events and focus on the more serious ones (as defined by the FAA), the coefficients become <u>more</u> rather than less negative. Taken literally, the data suggest that a passenger would have reduced his airborne death risk over this period by preferring mishap-prone airlines.

TABLE 1: Correlation Across Major Us Jet Airlines Between Nonfatal Accident/Incident Rates And Passenger Death Risk, 1/1/90-3/3/1/96

Type of Mishap	Correlation of Rate with Passenger Death Risk per Flight
Incidents Only	-.10
Incidents + Accidents	-.21
Accidents Only	-.29
Serious Accidents Only	-.34

NOTES: Incidents are events defined and reported as such by NTSB; accidents and serious accidents are as defined and reported by NTSB/FAA. Statistics shown are the coefficients of correlation between the mishap rate per 100,000 departures and the death risk per randomly-chosen nonstop flight (i.e. the Q-statistic).

One can devise an intuitive explanation for this perverse relationship. Suppose that all airlines suffer emergencies at the same rate, but that some are more adept than others atresolving them without dire consequences. Then the more skillful airlines will have relatively few disasters but relatively many nonfatal mishaps, while the less-adept carriers will exhibit the opposite pattern. It is a matter of elementary algebra to demonstrate that, in this situation, the coefficients of correlation would follow the general pattern of Table 1. Having said this, however, the authors are more inclined to treat the negative relationship as a coincidence tied to the paucity of disasters than a substantive phenomenon.

Data about all mishaps--whether they led to casualties or not--are exceedingly important to aviation professionals, who must learn whatever they can from all such events to “nip” potential tragedies in the bud. But statistics about nonfatal events are scarcely helpful in making comparisons among airlines, for such data bear no apparent positive relationship to passenger mortality risk. Using such data to assess either absolute or relative risks is devoid of both conceptual and empirical support. If broad listings of mishaps avoid the small-sample problems associated with fatal events, perhaps this is a situation where the cure is worse than the disease.

III. THE ESTABLISHED DOMESTIC TRUNKLINES

Both Barnett, Abraham, and Schimmel (1979; hereafter BAS) and Barnett and Higgins (1989; hereafter BH) paid great attention to the established domestic trunklines, namely, the US jet carriers that offered extensive interstate service prior to the Airline Deregulation Act of 1978. There were sixteen such airlines when BAS wrote; by 1996, however, mergers, bankruptcies, and closures had reduced their number to seven:

American Continental Delta Northwest
TWA United US Air(ways)

Over 1987-96, these carriers collectively performed 45 million nonstop flights. Fifteen of these flights suffered passenger fatalities, with death rates ranging from 1% to 100%. V, the number of full-crash equivalents these airlines experienced (i.e. the sum of x_i’s in the numerator of Q in (2)), was 6.63. Their overall Q-value was therefore 6.63/(45 million), which works out to roughly 1 in 7 million.

Given this Q-value, the number of nonstop flights a passenger could take before perishing in a fatal crash would follow a geometric probability distribution with parameter 1 in 7 million. Hence, a traveler who took one jet flight every day would on average go 19,000 years before succumbing to a fatal crash. This statistic means that, not only is the risk infinitesimal on individual flights, but even frequent flyers face minimal cumulative risk over their careers.

Of course, 1 in 7 million is the average death risk per trunkline flight, which raises the question whether the statistic varies from one trunkline to another. Before considering that issue, a brief digression might be helpful.

Digression

In its 1/28/98 issue, TIME magazine noted that the number 828 had come up in the Connecticut State Lottery on both 1/9/98 and 1/10/98. Emphasizing the rarity of such a coincidence, TIME described the outcome as a “one in a million” event. The magazine was correct in that the probability of getting (the three-digit number) 828 on these two days was $(1/1000)^2$ = 1 in 1 million. But it pays to ponder what was the real event that TIME considered noteworthy.

It seems likely that TIME would have thought the situation just as intriguing had the number 336 come up on the two days, or had the number 459 had come up on 1/4/98 and 1/5/98. There are thus many thousands of ways of getting the requisite “one in a million” event in Connecticut; indeed, the probability is 31% of getting at least one such coincidence there over the course of a year. And TIME would presumably have been just as impressed had a similar event occurred in Massachusetts, New Jersey, or California. The upshot is that, far from being rare, what TIME depicted as a singular coincidence was foreordained to occur somewhere in the US every month or so.

Is US Air Less Safe?

Table 2 presents data about the records of the seven domestic trunklines over 1987-96:

TABLE 2: Numbers Of Full-Crash Equivalents And Of Flights For Seven Domestic Trunklines, 1987-96:

Airline	Number of Domestic Full-Crash Equivalents	Number of Flights (Millions)
American	0	7.2
Continental	0.32	4.5
Delta	0.16	8.5
Northwest	1.21	4.9
TWA	0	2.7
United	1.40	6.5
US Air	3.53	8.6
TOTAL	6.63*	44.7*

NOTES: Scheduled domestic jet flights only. Includes flights by wholly-owned subsidiaries of these carriers that were subsequently absorbed into them. *Includes flights (and 0.01 full-crash equivalents) on established trunklines that ceased operations over 1987-96 (e.g. Eastern).

These statistics imply that US Air--which operated 19% of trunkline flights over 1987-96--amassed more full-crash equivalents than the other six carriers combined. Its Q-statistic of 1 in 2.5 million (3.53/8.6m) was about five times the statistic for the remaining airlines (3.1/36.1m). Does this difference reflect a temporary spasm of bad luck for US Air, or instead something systematic?

There is no immediate answer to that question, but the laws of probability offer some insight. If the trunklines were all equally safe, then, under an equal-safety hypothesis, the prior probability that US Air would have suffered any given one of the 15 fatal events on the trunklines would be .19 (i.e. its share of trunkline flights). We can therefore work out the distribution of US Air's V-value (full-crash equivalents) under an equal-safety hypothesis. Its probability that V=0, for example, would be $(.81)^{15}=.04$, given an 81% chance that it would avoid any particular fatal event. Had all x_i's over the period been either 0 or 1, the distribution of V would have been binomial; the actual distribution is a bit more complicated, with the possibilities ranging from 0 to 6.63.

At 3.53, US Air's V-value was at the 98.4th percentile of its distribution, meaning that, under a random distribution of crashes, the chance was only 1 in 60 that US Air that it would have sustained such a lopsided share of fatalities as it actually did. However, we should remember the Connecticut lottery and raise the question: assuming equal-safety, what is the chance that <u>one</u> of the seven trunklines would have been at least as high in its V-distribution as US Air was? Here the answer is roughly 7x .016, or about 11%.

In other words, there is a 1 in 9 chance that the worst loser in a random "lottery" of trunkline crashes would have fared as badly as US Air did (if not worse). Under usual standards, therefore, we have no statistically significant evidence that US Air was operating less safely over 1987-96 than other trunklines. It is interesting to note that US Air's Q-value--which was a factor of five higher than that for other trunklines over 1987-96--was roughly a factor of five lower than its competitors' statistic for 1977-86. In retrospect, it would have been unwise to assert on 1/1/87 that, given US Air's superb recent record, the carrier was unusually safe to fly over the next decade. Barring other information, making the opposite assertion because of US Air's subsequent record would seem equally unwise.

<u>A Test of the "Equal Safety" Hypothesis</u>

More formally, an equal-safety hypothesis within a group of airlines could be tested as follows:

(i) Work out the probability distribution for each carrier's number of full- crash equivalents, assuming that its prior probability of suffering each fatal event is equal to its proportion of flights.

(ii) Find the probability under that distribution that the carrier's number of full-crash equivalents would be at least as large as its observed number (i.e. compute the p-value of its result).

(iii) Find p_s, the smallest p-value among those computed (i.e. the one corresponding to the individual outcome most "hostile" to the equal-safety hypothesis). If there are M airlines in the comparison, then compute:

$$r = 1 - (1-p_s)^M \qquad (3)$$

The quantity r is the approximate probability that the worst of the M observed outcomes would be as "hostile" to the hypothesis as the outcome that actually arose. Thus, r is the overall p-value of the result under the worst-outcome test.

(iv) If $r > .05$, then one would not reject the equal-safety hypothesis under traditional standards. This means that the outcome can be viewed as reasonably consistent with the hypothesis (which, of course, falls short of a demonstration that the hypothesis must be true).

This "order statistic" test focuses only on the worst outcome because, given the scarcity of crashes, the best observed record is likely to be perfect whether equal-safety is true or not. Thus, the best outcome says little about whether the equal-safety hypothesis is viable.

IV. OTHER US JET OPERATIONS

Beyond domestic trunkline flights, the scheduled jet operations of US air carriers fall into three categories:

Flights of Established Regional Airlines

Established regional carriers offered extensive jet service prior to US airline deregulation, but such service was limited to particular regions of the country. Some of these airlines--which include Alaska, Hawaiian, Southwest, and Air California--greatly expanded their operations after deregulation, while others were merged into trunkline carriers. Whatever its literal appropriateness today, we refer to the survivors among these carriers as "regional."

Flights on "New Entrant" Carriers

"New entrant" carriers offered virtually no scheduled domestic jet flights before deregulation, but took advantage of the policy shift to start such service after 1978. Examples of such carriers are People Express, Tower Air, Air South, and Western Pacific. The heavy majority of these airlines--including all but the second of those just listed-- have ceased independent operations.

Scheduled US International Jet Flights

These flights originate and/or terminate outside the US domestic system (i.e. the 50 states plus Puerto Rico, the Virgin Islands, and Guam). By 1996, most US international flights were operated by the seven surviving trunklines; years ago, however, Pan Am was by far the leading US flag carrier.

Table 3 reports mortality risk for flights in all four categories of US jet operations, in both the last decade and the preceding one. Among other things, it reveals that the trunkline Q-value for 1987-96 which we have discussed was actually 50% higher than the corresponding statistic for 1977-86 (1 in 7m vs. 1 in 10.5m). For the full twenty-year period, we see polarization in Q-values between, on the one hand, flights on established US domestic carriers (trunkline plus regional), and, on the other, flights on new entrant carriers and on international routes. Death risk statistics for the latter operations are about a factor-of-six higher than those for the former.

Table 3: Death Risk per Flight for Four Categories of Scheduled US Jet Flights and Two Successive Decades

Category	1977-86	1987-96	(1977-96)
Trunkline	1 in 10.5m	1 in 7m	(1 in 8m)
Established Regional	1 in 3m	0	(1 in 10m)
New Entrants	1 in 1m	1 in 3m	(1 in 1.5m)
International	1 in 1.5m	1 in 1.5m	(1 in 1.5m)

m = million
Numbers in parenthesis are risk levels for 1977-96

However, before taking any of these outcomes at face value, we should inquire about their statistical significance. We first explore the time-trends in individual groups of flights; then we consider the differences in observed records across groups.

Time Trends

To assess whether mortality risk had shifted significantly over time, BH dichotomized fatal events into (i)disasters that killed at least half the passengers on board and (ii) other accidents that the majority on board survived. Then they explored whether the spread of disasters over two successive periods was consistent with an unchanging risk over time. If, for example, an equal number of flights were performed in both periods and there were Z disasters in total, then--absent any time trend--the number of disasters in the first period should have the same probabilistic behavior as the number of heads when a fair coin is tossed Z times. If the numbers of "heads" (or tails) among the crashes would seem highly unusual for a fair coin, then the hypothesis of stability over time would be rejected.

However, this test procedure has some unattractive features. Under it, fatal events with 1% and 49% death rates as treated as equivalent, as are events with rates of 51% and 100%. Rates of 49% and 51%, however, get strikingly different treatment. Moreover, the analysis considered not the specific date of a fatal event, but only whether it occurred in the earlier or later period. We thought it desirable to try to improve on such a procedure.

For the present study, we worked with a modification of the Kolmogorov-Smirnow statistical test. For the period from 0 to T and an intermediate time t, we define the following quantities:

----

0 t T

H(t) = fraction of all relevant flights over (0,T) that were performed during (0,t)

V(t) = fraction of allthe full-crash equivalents accumulated over (0,T)
that arose over the shorter period (0,t)

$D(t) = |(H(t)-V(t))|$

$G = \max\{D(t)\}$ for any t in (0,T)

Under the null hypothesis H_1 that prior death risk per flight was constant over (0,T), one would expect that H(t) and V(t) would be fairly close together at any time t. When 25% of the full period's flights have been completed, for example, we would anticipate under H_1 that roughly 25% of the full-crash equivalents would have arisen. Thus, neither D(t) nor G should stray "too far" from zero. The only reason these quantities might stray from zero at all is that, even if fatal events are distributed randomly over a period, some bunching can arise by chance alone, as can some intervals in which events are scarce.

To test H_1, we might first compute G using the actual data for a period. Because each flight generates a "proportion-killed" statistic that varies continuously between 0 and 1, we cannot use standard Kolmogorov-Smirnov tables to interpret the observed G-value. What we can do,
however, is:

(i) imagine that the casualty pattern observed (same number of fatal events with same passenger death rates) had arisen completely at random among the period's flights,

(ii) use Monte-Carlo methods with a random-number generator to determine the probability distribution of the statistic G assuming such a random spread of events, and

(iii) find the probability under this distribution of getting an observed G-value at least as high as the one that actually came up (i.e. work out the p-value of the observed outcome). Based on that p-value, we can make an assessment of the plausibility of H_1.

To test H_1 for the trunklines over 1977-96, we note that there were 24 fatal events among US domestic trunkline flights over 1977-96, with varying death rates. Given the exact dates of these events and the number of domestic trunkline flights in each of the twenty years considered, the actual G-statistic for the period was 0.13. (This value arose because, on 8/1/87, the carriers had performed 49% of their total flights over 1/1/77-12/31/96, but had only experienced 36% of their full-crash equivalents. At no other time during the two decades did the difference between the two percentages exceed this 17-point spread.) A computer was asked randomly to spread the 24 crashes with their particular death rates among the 86 million trunkline flights over the twenty years; in many thousands of such simulations, dates of these "crashes" were noted and G-values calculated. The Monte-Carlo simulation showed that, if crashes were randomly distributed (i.e. if H_1 were true and time-trends were absent), the calculated G would reach .13 or higher <u>fully 78% of the time</u>. Thus, under usual standards and the testing procedure we used, H_1 would not be rejected: the recent rise in trunkline death-risk in Table 2 could readily be construed as a statistical fluctuation rather than evidence of a meaningful decline in safety.

Indeed, the raw data provide powerful signals that the trunkline "time-trend" in Table Q should not be taken seriously. It turns out that nearly 25% of trunkline full-crash equivalents for the two decades 1977-96 arose in the single year 1987. Had we broken the twenty-year period into the two parts 1977-87 and 1988-96, therefore, the Q-statistic for the first period would be higher than for the second (1 in 7m vs. 1 in 10m), and the evidence of an adverse time-trend would disappear. An advantage of working with the G-statistic is that it never divides the full period into two distinct parts, so we need not worry about the sensitivity of the result to the break-point chosen.

Not only does the G-test work against a time-trend for the domestic trunklines, but it also does so for the other three groups of jet operations. Table 2 hints that the established regional carriers improved over time, but they suffered exactly one fatal crash (which had no survivors) over 1977-96, which makes it hard to speak seriously of any time-trend. The data about new-entrant and international flights were likewise consistent with temporal stability. In sum, there is no clear reason to argue that US jet safety either improved or got worse in the decade 1987-96 compared to the previousone.

Cross-Group Risk Comparison

To explore the equal-safety hypothesis (hereafter H_o) for the four groups of flights, we could apply the same "worst case" method as for the trunklines, based on the probability that--under a random distribution of fatal events--the highest observed p-value for the four groups would have been as high as the value that came up. If we work with the twenty-year data in Table 2, the highest p-value was for the international flights. (The new-entrants had the same 20-year Q-value as the internationals in Table 2 but, because there were far more international than new-entrant flights, the outcome for the former is more statistically "freakish" under an equal-safety hypothesis than that for the latter.) The overall test statistic (i.e. r in (3)) when all four groups are considered follows r= 0.04. This result implies that, assuming that US jet crashes were distributed randomly, the chance is 4% that the "least lucky" of the four flight groups would have experienced as disproportionate a share of full-crash equivalents as international flights sustained over the study period.

If, however, the analysis is limited to the three groups that operate in the US domestic environment, then the new entrant carriers had the worst observed outcome and the test-statistic for Ho becomes r= .07. This p-value is higher than that which BH reached in a comparison between new-entrant jet airlines and established trunklines for1979-86 (i.e. .02); hence, the "strength of evidence" against the new entrants is less for the first eighteen years after deregulation than it was after the first eight..

In both the four-group and three-group comparisons, therefore, the observed variation in Q-values over 1977-96 falls near the "borderline" of statistical significance at the usual 5% level . But statistical significance should not be interpreted in a mechanical way: some "significant" patterns are neither important nor informative, which other patterns that fail to attain significance are not illusory. We will say more about the international outcome Later; now we turn to a more detailed review of the new-entrants' recent safety record.

The Valujet Crash

The new entrants had only one fatal experience over 1987-96, but that event--the 1996 crash of Valujet Flight 592 into the Everglades--was among the most influential disasters in American aviation history. It solidified the perception that new-entrant carriers were less safe than established ones, and contributed to the cessation of scheduled service on Air South, Western Pacific, Nation's Air, and Carnival. Valujet itself changed its name to Airtran because, as one of its officials explained, "you're known for what you're known for."

From a statistical standpoint, the market response to the disaster may have been an overreaction. Valujet completed 150,000 flights over 1987-96 (it began operations in 1993); given passenger-fatality patterns for established carriers, the probability was .015 that its observed Q-value would have been as high as it actually was (namely, 1 in 150,000). This probability is similar to that we saw earlier for US Air and, in analogy with the trunkline-discussion, the probability was greater than10% --under established-carrier patterns—that some new-entrant carrier would have had as adverse an observed record for 1987-96 as Valujet did.

Furthermore, it makes sense to consider the circumstances of the Valujet crash. If the plane had been felled by a meteorite, no rational person would have suggested that the event proved the lesser safety of the new entrants. What actually happened was that a maintenance contractor--which also worked for some of the trunkline carriers--sent Valujet some cargo that included oxygen canisters that were explicitly –but erroneously--described as nonhazardous and "empty." Adam Bryant, a key aviation correspondent, wrote in *The New York Times*, that "(other airlines) know that the kind of human error that appears to have led to the Valujet crash-a mistake in labeling a box of hazardous oxygen generators that were put on board flight 592-could easily befall any one of them" (6/11/96; Section 4, page 1). An obvious question arises: if the event could "easily" have befallen the established trunklines, does it clearly demonstrate that Valujet was less safe than these carriers?

In the extraordinary month after the Valujet crash--which ended with the FAA's grounding the airline for over three months--it was repeatedly noted that, prior to the Everglades disaster, Valujet had sustained an unusually high rate of nonfatal accidents and incidents. Mary Schiavo, Inspector General of the Department of Transportation, has written that, because of this pattern, Valujet was "primed for a major crash." (Schiavo [1997], at page 7). We saw in Section II, however, that there is no evidence independent of Valujet that higher involvement in nonfatal events suggests a greater propensity to suffer disasters.

In short, the evidence against the new entrants from the last decade studied consists of one crash that, arguably, was not the fault of the carrier that suffered it. Whether the post-deregulation carriers were at higher risk in recent years--let alone whether they present higher passenger mortality risk over the years ahead--is thus questionable.

AUTHORS' NOTE: One of the participants in this study--Barnett--was hired by Valujet's law firm after the crash of Flight 592 to assist the company in its submissions to Congress, the US Department of Transportation, and the federal courts. His written analysis in that setting was similar to the one presented here.

V. US COMMUTER SERVICE

Short-haul commuter flights performed by propeller or propjet aircraft have become increasingly important in the US domestic air system. Over the last decade, there were 60% as many commuter flights as jet flights; most commuter operations were conducted by airlines either affiliated with or wholly-owned by established trunklines. The safety of commuter service--especially relative to jet service--has been a subject of much discussion, little of it favorable. Concern reached a peak in late 1994 after two American Eagle planes crashed in quick succession.

Neither BAS nor BH considered commuter air safety, but there are no conceptual or practical impediments to doing so. Over 1987-96, US commuter operators performed approximately 35 million flights and suffered 19.08 full- crash equivalents, which yield a Q-value for the group of approximately 1 in 2 million. Table 4 presents mortality risk information for eight of the largest commuter carriers; our test of the equal-safety hypothesis Ho yields no evidence of statistically significant disparities in passenger risk (r =.8).

TABLE 4: Death Risk per US Commuter Flight on Eight US Carriers, 1987-96

Airline	Death Risk per Flight (Q-value)
Alaska Airlines Commuter	0
American Eagle	1 in 1.5 m
Continental Express	1 in 1.5 m
Delta Connection	1 in 1.5 m
Northwest Airlink	1 in 1.5 m
Trans World Express	0
US Air Express	1 in 10 m
United Express	1 in 1.5 m
TOTAL	1 in 2m

NOTES: Some approximations were used in the calculations; denominators of risk estimates are rounded off to nearest half million.

More complicated is the question of how the commuter carriers compare with the jet airlines. The commuters' Q-value of 1 in 2 million is higher than those for 1987-96 on domestic jet carriers; it was not higher, however, than that for US international jet flights. If one compares the collective record of domestic jet services(trunklines, regionals, and new entrants) with that of the commuter airlines, one finds that the Q-value for domestic jets was 1 in 7 million, and that the higher observed risk for commuter passengers was statistically significant (r=.01).

However, the logic behind the last comparison is not airtight. In the previous section, we excluded international jet flights from an analysis of US jet safety because they operate in different flying conditions than domestic flights. But US commuter operators can also face different conditions from US jets. The last commuter crash during the study period, for example, was a 1996 runway collision at a rural airport without a control tower. The question of commuter air safety, therefore, may be entangled with other issues that make a head-to-head contrast with the jets a questionable exercise.

In any case, the jet/commuter question is somewhat moot because there are few routes on which the two types of airlines compete. Because the car is the real competitor to the commuter plane, the interesting issue is how commuter flights compare with auto trips of similar distances. Researchers have estimated that the kinds of drivers who might use commuter flights--sober, seat-belted, over 40 years old and in a heavier-than-average car--suffer a death risk of approximately 1 in 4 million on an intercity trip of 300 miles (Evans, Frick, and Schwing [1990]). That risk estimate may be too low (see Barnett (1991)),but it is probably fair to say that commuter flights pose about the same death risk as the auto trips they replace, and not--as is sometimes implied--a far lesser risk. On the other hand, the chance of a nonfatal injury is considerably higher in the car than on the propjet.

VI. NON-US FIRST-WORLD JET OPERATIONS

BH designated 21 nations besides the US as constituting the economically and technologically advanced First-World (e.g. Australia, Belgium, Canada, Finland, Israel, Japan), and proceeded to investigate mortality risk on the scheduled international flights of each country's primary flag carrier (e.g. Air France). Here we substantially generalize the BH analysis: we consider all scheduled domestic and international jet services, rather than none of the former and some of the latter. And--last but not least--we add New Zealand to the First-World list.

Table 5 presents Q-values for the last two decades for First-World jet services outside the US. We see that death risk fell dramatically on domestic flights--from 1 in 2 million to 1 in 11 million--but remained stable on international routes. 85% of these domestic full-crash equivalents over 1977-96 had already occurred by October 1985, at a time when only 42% of the period's flights had been performed (p-value=.01 in modified Kolmogorov-Smirnov test of H_1; however, the chance of getting one such "false positive" in this paper's eight time-trend analyses is about .08.). Tests of the equal-safety hypothesis H_0 across the domestic or international airlines studied provided no statistically-significant basis for identifying any as especially dangerous.

TABLE 5: Death Risk Per Flight In Scheduled First-World Operations For Non-US Airlines, 1977-96*

Domestic:

Period / Category	Death Risk Per Flight
1977-86	1 in 2 million
1987-96	1 in 11 million

International:

Period / Category	Death Risk Per Flight
1976-86*	1 in 5 million
1987-96	1 in 4 million

* Initial period is 1976-86 because that is the period considered in BH.
NOTES: Sources: Some approximations were used in the calculation; see Appendix.

Air Supremacy Lost?

In August, 1996, President Clinton described US airlines as “the safest in the world” (a characterization also made by many others). BH had indicated that, over the quarter-century ending in 1986, US domestic trunklines outperformed all other groups of carriers studied to a statistically-significant extent. Yet, Table 6--which merges data from Tables 3 and 5--suggests that statistics from 1987-96 offer a different picture.

TABLE 6: Death Risk per Scheduled Jet Flight over 1987-96 for US and non-US First-World Airlines

	Domestic	International
US	1 in 7 million	1 in 1.5 million
Non-US	1 in 11 million	1 in 4 million

If the assertion that US carriers are “safest” means that other First-World nations need US help in keeping their planes aloft, it is apparently out of date. However, as Table 7 suggests, we need not go to the other extreme and lament the decline and fall of US aviation safety. The strong recent safety gains of non-US domestic jet carriers mirror the pattern among US domestic trunklines ten years earlier. And, as we have noted, the faltering of the Q-value for US trunklines over 1987-96 could easily be a fluctuation, and was based almost wholly on events in 1987. Indeed, if we consider the period 1988-97, the Q-statistic was 1 in 10 million both for all US domestic jet carriers (trunklines, regionals, and new entrants combined) and all non-US domestic jet carriers.

TABLE 7: Follow The Leader?: Death Risk Per Domestic Jet Flight In Two Twenty Year-Periods And Two Airline Groups

US Domestic Trunklines:

1967-76	1 in 2 million
1977-86	1 in 11 million

First-World Domestic Jet Carriers Outside US:

1977-86	1 in 2 million
1987-96	1 in 11 million

NOTE: Denominators in death-risk statistics are rounded off to nearest million.

US jet carriers can still be considered the safest in the world in that they stand comfortably among those airlines that have been most successful in preventing passenger deaths. But they are no longer manifestly superior to these other airlines, any more than the US is manifestly superior to the rest of the First-World in auto safety, rail safety, industrial safety, or overall life expectancy.

Table 8 pools the two columns of data in Table 6 to get overall statistics for the First-World:

TABLE 8: Death Risk Per Scheduled Jet Flight Over 1987-96 For All First-World Airlines, In And Outside The US

Type of Flight	Q-Value
Domestic	1 in 8 million
International	1 in 3 million

VII. JET OPERATIONS IN THE DEVELOPING WORLD

Having defined all nations we did not classify as First-World as part of the Developing World, we proceeded to estimate mortality risks for domestic and international jet travel aboard airlines in developing countries. Gathering data about these operations was a formidable task; the problems were sufficiently acute for the former Soviet Union that we decided to exclude it from the analysis.

For about 30% of fatal jet events in developing nations, our usual data sources did not show whether the affected flight was on a domestic or international trip. We contacted Boeing's Product Safety office and learned that, among the Boeing jets within this group, 77% of those with known routings were performing domestic flights. Even Boeing, however, was unaware of the routings of 30% of these flights. As a best guess, we assigned 77% of the jet full-crash equivalents among affected flights to domestic operations, and 23% to international services.

Table 9 summarizes Developing-World Q-values for the last two decades. The risk number are far larger than their First-World counterparts, and there is no evidence of reduced mortality risk during the last decade. Contrasting Table 9 with Table 8, we see that death risk per domestic jet flight over 1987-96 was approximately 15 times as high in the Developing world as in the First World. Death risk on international flights was a factor of 7 higher. These First World / Developing World differences are of strong statistical significance and, if the more limited data in BAS and BH are any guide, there is nothing new about such large disparities.

TABLE 9: Death Risk Per Scheduled Jet Flight In The Developing World, For Domestic And International Operations Over The Two Decades 1977-96:

Type of Flight	Q-values:	
	1977-86	1987-96
Domestic	1 in 500,000	1 in 500,000
International	1 in 600,000	1 in 400,000

NOTES: See Appendices for more details of these calculations.

Of course, the Developing World is far from perfectly homogeneous, and many individual airlines situated there that had no fatalities over 1987-96 (or even 1977-96). But this circumstance does not invalidate an equal-safety hypothesis: under a random distribution of crashes, many airlines would have been expected to have Q-values of 0. To explore the issue further, we might focus on those developing nations for which recent political and / or economic changes may have moved them closer to First-World status: especially the "Asian Tigers," and the former Soviet satellites in Eastern Europe.

As Table 10 suggests, the 1987-96 Q-values for these countries are far closer to the overall average for the Developing World than the First World. It therefore appears that, while it is implausible that all Developing World airlines are absolutely identical in safety, there is no obvious partitioning rule to distinguish the less- safe from the more-safe.

TABLE 10: Death Risk Per Jet Flight For Airlines From The "Asian Tigers" And From Eastern Europe, 1987-96

Carrier Group	Full-Crash Equivalents	Number of Flights(Millions)	Q-value
Asian Tigers	5.94	4.2	1 in 700,000
Eastern Europe	1.68	0.7	1 in 400,000

NOTES: Statistics are based on all jet flights, both domestic and international. Asian Tigers are defined as South Korea, Singapore, Hong Kong, Taiwan, Malaysia, and Thailand. Eastern European countries considered are Poland, Hungary, Rumania, Czech Republic/Slovakia, and Bulgaria. (Albania has no jet airlines, and the former Yugoslavia was excluded because of conflicts.)

An Ecological Fallacy?

Consider the proposed syllogism: John is better in math than Bill; I have a problem in trigonometry; therefore, I should go to John. The flaw in this reasoning is that, while John might be stronger overall in math, trigonometry might be his weak point and Bill's strong point. The assumption that a general statement is true in every particular case is known as the ecological fallacy.

The relevance of the fallacy here is that, while First-World carriers have far lower Q-statistics than Developing-World ones, one cannot simply assume that such a difference prevails on routes served by both kinds of airlines (e.g. Paris-Karachi, Tokyo-Delhi, Miami-Caracas). To investigate the issue, one must calculate relevant Q-values for the subset of international routes between First-World and Developing-World cities. Table 11 does so for Developing-World and First-World carriers over 1987-96.

TABLE 11: Death Risk Per Scheduled Jet Flight Between The First World And The Developing World For Two Groups Of Airlines, 1987-96:

Carrier Group	Full-Crash Equivalents	Estimated Number of Flights (millions)	Q-value
First-World	4.03	2.5	1 in 600,000
Developing World	4.98	4	1 in 800,000

NOTES: Estimated numbers of flights subject to sampling error; see Appendix. Differences in observed Q-values for the two groups are not statistically significant. Classification of airlines as First-World or Developing World is based on their home offices.

We see that the First-World safety advantage disappears on these “hybrid” routes, and that First-World airlines suffer mortality risks in this setting much closer to the norms for the Developing World than to those for their home countries. We should avoid glib speculation about the meaning of Table 11: some of the listed crashes had no apparent relation to the First World / Developing World dichotomy (e.g. the thrust reversers deployed in flight on one First-World Boeing-767, which thereupon crashed in Thailand). But the Table hints that the relatively-difficult flying environments in Developing World countries pose hazards to all carriers that fly there, regardless of their national origins.

VIII. SUMMARY AND CONCLUSIONS

We have analyzed statistics about mortality risk in recent scheduled air travel. While such data offer a useful synopsis of “where we stand,” we have said little about why aviation accidents occur, or what should be done to reduce them in the future. That formidable topic should be the subject of other papers.

The main conclusions we have reached are these:

* When measuring passenger mortality risk, one should worry about statistical indicators that pay no attention to the survival rates on fatal crashes, or that tacitly assume that the risk of a flight is proportional to its length or duration. One need carefully consider the statistical volatility in data about rare air crashes, but should avoid either exaggerating the consequences of such volatility or ignoring them altogether.

(In this paper, we have used a mortality-risk measure that tried to circumvent the hazards above, and have proposed tests for the statistical significance of observed patterns that supplant more rudimentary tests in the two previous MIT studies.)

* One should beware of the use of data about nonfatal accidents and incidents to “flesh out” an airline’s safety record, and of the assertion that such data offer a more stable proxy for its underlying death risk than its recent fatalities record. There is, if anything, a negative correlation between involvement in such nonfatal events and involvement in deadly crashes.

* Passenger death risk per US domestic jet flight was about 1 in 7 million over the decade 1987-96. That statistic does not represent an improvement over the previous decade, but it does imply that, if a passenger took one domestic jet flight at random every day, she could on average travel 19,000 years before succumbing to a fatal crash. While US Airways had more full-crash equivalents over 1987-96 than all other established carriers combined, the pattern could well reflect bad luck rather than any systematic factor.

* The “new entrant” jet carriers formed in the US after airline deregulation suffered only one fatal event over 1987-96, but it was the notorious Valujet crash in 1996. Statistical evidence fails to establish that Valujet was operating less safely in recent

years than major US jet carriers, or that other new entrants--which had no fatalities over the decade--were doing so.

* The death risk per flight on US commuter flights was about 1 in 2 million over 1987-96. This excess over the corresponding rate for domestic jets was statistically significant. Although it is a truism that planes are far safer than cars, careful drivers who replace commuter flights with auto trips do not thereby heighten their mortality risk.

* In First-World countries other than the US, death risk per domestic jet flight fell from 1 in 2 million over 1977-86 to 1 in 11 million over 1987-96. The corresponding risk on international flights was stable at about 1 in 4 million (as opposed to a stable risk of 1 in 1.5 million on the international jet flights of US carriers).

* It is no longer accurate to proclaim US jet carriers "the safest in the world," for, over 1987-96, their counterparts in the rest of the First World achieved lower death risk than they did on both domestic and international operations. However, this reversal is not statistically significant; indeed, if the focus is on 1988-97 rather than 1987-96, the two groups of domestic carriers have nearly identical death risk (1 in 10 million per flight). It appears that First-World nations are quite homogenous in aviation safety, much as they are on most other dimensions of mortality risk.

* Developing world airlines continue to be generally riskier than those in the First-World: compared to First-World levels, their mortality risk for 1987-96 was roughly a factor of 15 higher for domestic jet operations. However, the discrepancy was smaller for international flights and, for those international routings on which First-World and Developing World airlines compete, there was virtually no difference is mortality risk over 1987-96, with each group achieving a death risk per flight of roughly 1 in 700,000.

Nowhere to Run?

If we stand back from the specific findings of this study, two major points become apparent. The first is that, after decades of steady worldwide declines in airline death risk, the period 1987-96 saw stability rather than improvement. (The conspicuous exception is for domestic First-World jet operations outside the US.) The worldwide average risk level at which such stability occurred does not seem the lowest achievable, given that certain groups of airlines regularly do far better.

And, in an ironic outcome of an airline safety study, it appears that passengers have "nowhere to run": on competitive routes, there is rarely any reason related to safety to prefer one carrier over another. US domestic jets might be safer than commuter propjets, but the two groups of planes have different route networks: few jets fly 100-mile routes, and few propjets fly 1000-mile ones. Domestic jets flights may be substantially safer in the First than the Developing World; by definition, however, the two different route systems do not overlap at all. On routes between the First and Developing worlds, airlines with countries of origin in both worlds appear to fare about equally well.

In attempting to improve aviation safety, therefore, we should perhaps focus less on the achievements of particular airlines than on the hazards of different flying environments. Much as a rising tide lifts all boats, steps to improve the less-safe environments could benefit all airlines that fly through them and not just the local carriers.

ACKNOWLEDGEMENTS

Several individuals at FAA's Office of System Safety—especially Carolyn Edwards, Jack Wojciech, and Christopher Hart (Deputy Assistant Administrator for Aviation Safety) offered an abundance of encouragement and thoughtful, stimulating comments. Todd Curtis of the Boeing Company provided much information about Developing -World air crashes that we might not otherwise have found. Amedeo Odoni of MIT provided continuing intellectual and moral support. None of these people should be held responsible for any viewpoints expressed in this report.

REFERENCES AND PRIMARY DATA SOURCES

Barnett, A. “It’s Safer to Fly,” Risk Analysis, 11(1), 1991
Barnett, A., M. Abraham, and V. Schimmel, “Airline Safety: Some Empirical Findings,” Management Science, 25 (11), November 1979

Barnett, A. and M.K. Higgins, “Airline Safety: The Last Decade,” Management Science, 35(1), January 1989

L. Evans, M. Frick, and R. Schwing “Is it Safer to Fly or Drive?” , Risk Analysis,10(2), 1990

Flight International, annual air safety summaries in late January or early February issues, 1977-97.

Official Airline Guide, North American Edition, published by Reuben H. Donnelly, various issues, 1977-96

Official Airline Guide, Worldwide Edition, published by Reuben H. Donnelly ,various issues, 1977-96

Schiavo, M. Flying Blind, Flying Safe , Avon Books, New York (1997)
Statistical Summary of Commercial Jet Aircraft Accidents, Worldwide Operations 1959-97, Boeing Commerical Aircraft Company, 1997

Traffic, ICAO Digest of Statistics, Series T, various editions, 1976-96

APPENDIX A: Estimation of Death Risk per Flight

As noted in the text, the Q-statistic (death risk per flight) for a given setting follows the equation:

$$Q = V / N$$

where V= number of full-crash equivalents and N= number of flights

Our various estimates of V are based on annual summaries in *Flight International*, supplemented by records on various web pages (e.g tcurtis@airsafe.com). As noted, Boeing Product Safety provided data about whether certain fatal events occurred on domestic or international flights. In two instances, passengers died because of contaminated food (one person on British Airways, and one on Aerolineas Argentinas); we counted these events in the Q-calculations because they were apparently within the purview of the airlines involved. Though it is conceivable that some relevant accidents did not reach our attention, we have no reason to believe that such omissions are numerous. (A full listing of the events used in calculating V-values is available from abarnett@mit.edu; the key V-values themselves appear in Appendix B.)

We emphasize yet again the statistical volatility inherent in observed V-values. Suppose, for example, that an airline performed one million flights over 1987-96 and that—reflecting the carrier's long-term level of safety--each such flight independently had a 1 in 1 million probability of suffering passenger fatalities. (For simplicity, we assume there are no survivors on its ill-fated flights.) Familiar calculations reveal that, over 1987-96, this carrier would have a 37% chance of having no fatalities, a 37% chance of having one fatal crash, and a 26% of having two or more. Only in the middle case will the calculated Q-value reflect the true risk that passengers faced as they boarded their flights (1 in one million); in the other cases--which arise 63% of the time--the observed Q-statistic will differ from the underlying mortality risk by at least a factor of two. And, if two airlines shared this statistical profile (one million flights, 1 in 1 million death risk for each), the chances exceed 2 in 3 that their observed Q-values would be at least a factor of two apart.

Because the high "noise level" in individual V-values necessarily carries over to the Q-ratio itself, modest imprecision in estimating N (the number of flights completed) can be treated as a second-order effect. We took this point into account in approximating how many flights of various kinds were performed, not least because few exact numbers were readily available. While fairly precise records exist for scheduled US jet operations, there was no place we could go to get (say) the number of nonstop jet flights by First-World airlines between the First and Developing Worlds.

We often focused on the June 1992 issue of the *Official Airline Guide* (OAG; both Worldwide and North American editions), and on the weekly numbers of listed nonstop flights for the middle of that month. June 1992 is just past the middle of the period 1987-96; thus, whether a certain type of flight was becoming more or less frequent over the decade, the weekly rate for mid-June 1992 should approximate the mean weekly rate for the full decade. It is helpful that mid-June is neither a peak travel period nor at the depth of an off-season (although, given the costs of grounding an airplane, seasonal variations in demand should show up more in the fraction of seats occupied than in the number of flights performed). To illustrate how we used the June 1992 schedules, we describe two examples.

As one part of our effort, we needed to estimate the number of domestic jet flights in Spain for 1987-96. We first found the weekly number of nonstop domestic flights out of Madrid for mid-June 1992; then we doubled that number to take account of flights into Madrid. Thereafter we did the same for Barcelona, with the exception that we excluded all nonstop flights to and from Madrid. Then we considered Seville, excluding nonstop flights to/from Madrid and Barcelona. We proceeded in this way for any Spanish city that had nonstop service from any city we had previously considered, stopping when we ceased to find additional flights.

We took the total number of weekly nonstop flights thus uncovered and multiplied by 50 to approximate an annual rate. Then we multiplied by 10 to cover the decade 1987-96. We estimated that, for Spain over 1987-96, there were roughly 1.2 million scheduled domestic jet flights.

For flights between the First and the Developing World, we examined every tenth page in the OAG (all pages ending in 5), and counted every nonstop listing during a week in mid-June 1992. (We looked at 123 such pages.) On page 1195, for example, we found that First-World airlines operated 14 nonstop flights per week from Manila to Tokyo (JAL, Northwest), while Developing-World carriers operated 11 (Phillipine, Egyptair, Pakistan Int'l) The totals for each category were multiplied first by 10 (to reflect the full OAG rather than a 10% sample), and then by 50 and 10 to extrapolate from the one week studied to the full decade. As noted in the text, we estimated a total of roughly2.5 million "between-world" flights over 1987-96 by First-World carriers, and 4 million by Developing-World ones.

Obviously, these estimation procedures--though highly labor-intensive--are by no means exact. But the key question is whether the imprecision they introduce into the calculations compromise any of the conclusions drawn. We think not, given that no parameter estimates or comparative statistics are meant to be taken literally. We estimated, for instance, that mortality risk per domestic jet flight was roughly 15times as high over the decade studied in the Developing-World as in the First-World. Perhaps the exact ratio was 13 or 18, but would knowing of that revision be of paramount importance? When two estimated Q-statistics are close together, we have routinely warned that the difference could well reflect statistical fluctuations rather than anything meaningful. If we knew the exact numbers of flights completed, perhaps the ordering of the two observed Q-statistics would change. But we would be just as insistent in warning that the difference might be meaningless.

Heeding the advice that "the best is the enemy of the good," we have accepted some imperfections in estimating the numbers of flights performed. Without such compromises, we could not feasibly have completed certain calculations that had not, so far as we know, been performed in either the previous MIT studies or anywhere else. We remind readers that other problems in the available data (e.g. in whether certain Developing World crashes involved domestic or international flights) preclude claims of exactitude under any circumstances.

APPENDIX B:
Key Summary Statistics For Scheduled Nonstop Flights, 1987-96

Flight Grouping	Full-Crash Equivalents	Estimated Total Number of Flights (millions)	Death Risk per Flight (Q-value)
(1)US Domestic Jets	7.63	55	1 in 7m
(2) Other First-World Domestic Jets	1.65	18	1 in 11m
(3)Domestic Jets, Developing World	25.6	12	1 in 500,000
(4)US Domestic Commuter Propjets	19.08	35	1 in 2m
(5)All International Jets, US Carriers	3.00	4	1 in 1.5m
(6) All International Jets, Other First-World Carriers	4.06	16	1 in 4m
(7)All International Jets, Developing World	21.3	8.5	1 in 400,000
International Jets Between First and Developing Worlds:			
(8A) First-World Carriers	4.03	2.5	1 in 600,000
(8B) Developing-World Carriers	4.98	4	1 in 800,000

NOTES:
All statistics about numbers of flights performed involve some degree of approximation, though the degree is minimal for US jet services.

Overall numbers of full-crash equivalents on Developing-World airlines (in (3) and (7)) are subject to uncertainty of about 10%, because it is not clear whether certain jets that crashed were on domestic or international routes.

Because of statistical volatility in Q-values computed from small numbers of fatal crashes, we caution that the following differences in Q-values calculated above are not statistically significant: (1) vs. (2), (5) vs. (6), and (8A) vs. (8B).