Uses and Misuses of Risk Metrics in Air Transportation

CARMEN TERESA VILLARREAL

In air transportation the use of different statistics to measure risk (henceforth referred to as "risk metrics") yields different perceived levels of safety. The perception of risk, in turn, affects the framework within which thresholds for risk acceptability are set. Therefore, it is important to carefully consider the characteristics of several risk metrics before deciding to use a particular one. In this paper the strengths, limitations, uses, and potential misuses of seven commonly used risk metrics are examined. The risk metrics analyzed are (a) fatalities per hour of exposure to air transportation, (b) passenger fatalities per 100 million scheduled passenger-mi, (c) fatal accidents per 100,000 flights, (d) probability of being killed in an air carrier accident, (e) total accidents per 10 million system flying hours, (f) miles flown between successive accidents, and (g) mean time between failures. The conclusion is that there is no "unique" or "correct" way of measuring risk in air transportation. Therefore, air transportation risk studies should include a statement of the applicability, spectrum, and limitations of the chosen risk metrics.

The statistics used to measure risk (henceforth referred to as "risk metrics") play an essential role in evaluating the safety of the air transportation industry as well as in the comparison of the safety levels of different airlines, different modes of transportation, and different risk-generating activities. Risk metrics also play a crucial role in analyzing the effect of introducing changes in the air transportation system. For example, use of a particular metric may influence decisions such as whether to fly with two engines over the Atlantic or to increase the number of operations at an airport.

The use of different metrics yields different perceived levels of safety, which in turn affect the framework within which thresholds for risk acceptability are set. Because there is no "unique" or "correct" way of measuring risk in air transportation, it is important to explore the strengths and limitations of some frequently used risk metrics within the context of their use before deciding to use any one of them.

In this paper the following seven risk metrics are examined:

- Fatalities per hour of exposure to air transportation,
- Passenger fatalities per 100 million scheduled passenger-mi,
- Fatal accidents per 100,000 flights,
- Probability of being killed in an air carrier accident,
- Total accidents per 10 million system flying hours,
- Miles flown between successive accidents, and
- Mean time between failures.

These metrics were selected because they were considered representative of those that have been used and those that could be used in the air transportation industry.

The metrics studied in this paper have different mathematical interpretations and derivations, as well as different spectra of applicability. For example, four of the seven are built by dividing a mishap parameter (fatalities or accidents) by an activity parameter (hours of exposure, passenger miles, flights, or flying hours). Metrics built in this way acquire the shape of the parameter with the higher variance. Figures 1–3 show that the trend of the metric *passenger fatalities per passenger mile* closely resembles that for *passenger fatalities*, which is the high-variance parameter shown in Figure 1.

To examine the spectra of applicability of the aviation risk metrics, they are classified into three categories: wide spectrum, medium spectrum, and narrow spectrum. Wide-spectrum risk metrics are used to measure the risk posed by the air transportation system as a whole. Medium-spectrum risk metrics measure the risk posed by air transportation subsystems such as air traffic control, airline, airplane type, airplane maintenance schedule, and operational condition. Narrow-spectrum risk metrics measure the risk posed by characteristics or components of a subsystem such as human error, avionics, engines, and power plants.

DATA BASE

The data base contains the necessary parameters to build the metrics of interest. The actuarial data gathered reflect the accidents and fatalities of the U.S. domestic scheduled air carriers (Table 1) and the activity levels measured by passenger emplanements, passenger miles, load factors, aircraft miles flown, aircraft hours flown, and number of flights (Table 2). The data base covers the period from 1970 to 1985. U.S. domestic scheduled air carrier operations were chosen because they constitute a subgroup of the air transportation industry for which consistent data are readily available.

Military, nonscheduled carrier (typically represented by charter flights), cargo, general aviation, and international operations of U.S. air carriers were not considered in this analysis. The operations excluded constitute about 50 percent of the total U.S. air traffic activity (*21*). This exclusion does not limit the usefulness of the analysis, because the purpose of this paper is to illustrate the uses and misuses of different risk metrics in air transportation and not to evaluate the safety of the air transportation industry as a whole.

Boston University Law School, 13 Potter Pond, Lexington, Mass. 02173.

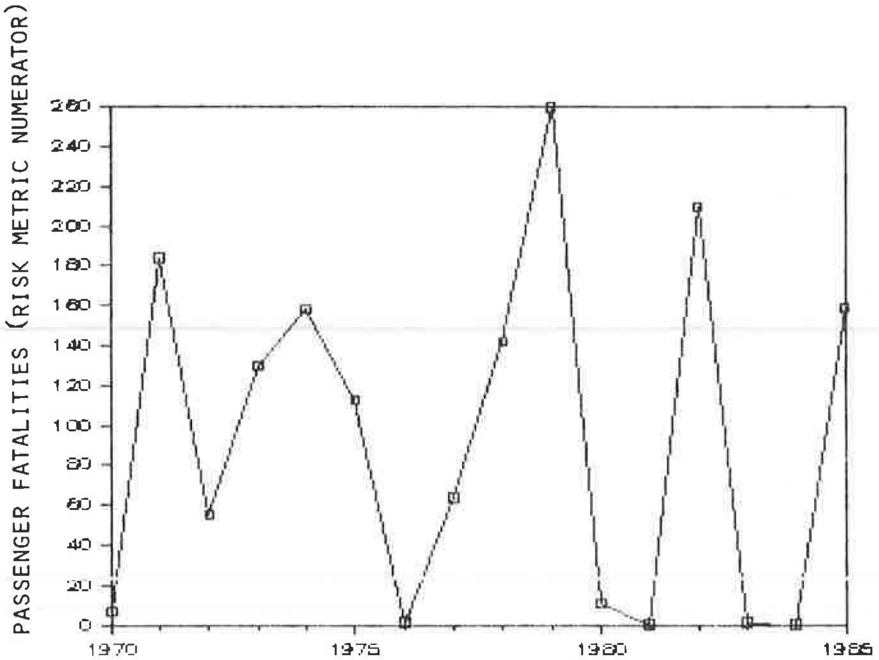

FIGURE 1 Composition of a risk metric: passenger fatalities.

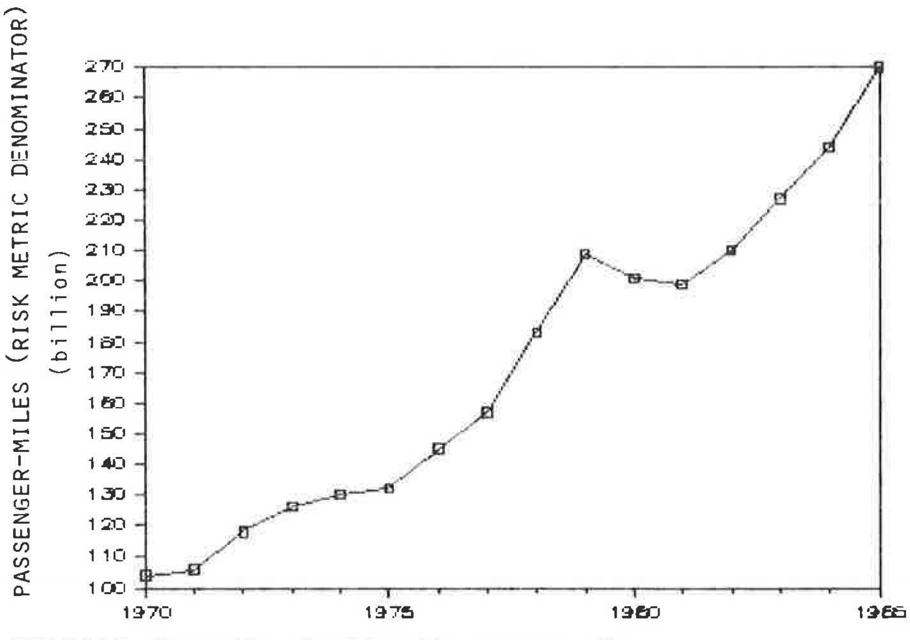

FIGURE 2 Composition of a risk metric: passenger miles.

USING ACTUARIAL DATA TO MEASURE RISK IN AIR TRANSPORTATION

Risk metrics derived from actuarial data can introduce two main types of bias: leveling of distinctions and omissions. "Leveling of distinctions" refers to the blending of data so that distinctions among individual data points are lost. Although this blending is the basis of statistical sampling, it can work against accurate reflection of the facts. An example is a metric that has fatalities in the numerator but does not distinguish among crew, passenger, and ground fatalities. This distinction is critical in many cases, for example, when the parameter of interest is willingness to undertake risk.

Omissions occur when relevant factors are not considered when a metric is built. For example, the metric *passenger fatalities per 100 million scheduled passenger-mi* considers only those passenger injuries that result in dealth within 7 days of an accident, omitting any accident-related casualties occurring more than 1 week after the accident.

Two additional problems arise from using actuarial data to measure risk in air transportation. One is that accidents are rare events; thus perceived variations in risk might be due to

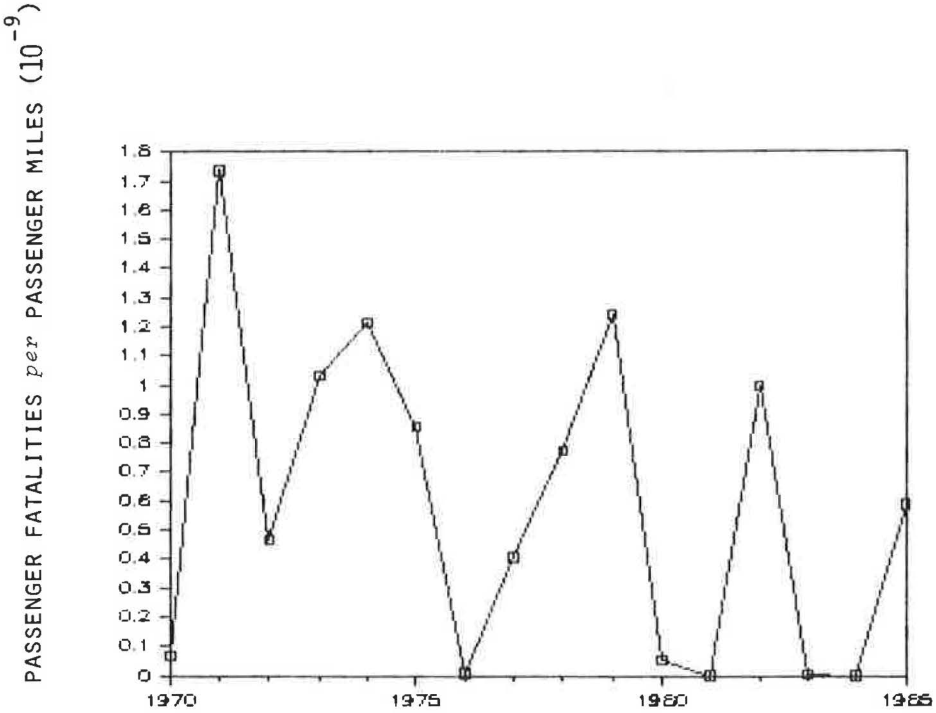

FIGURE 3 Composition of a risk metric: passenger fatalities per 100 million scheduled passenger-mi.

TABLE 1 ACCIDENTS AND FATALITIES OF THE U.S. DOMESTIC SCHEDULED AIR CARRIERS

	ACCIDENTS (1)		FATALITIES (2–17)			
	TOTAL	FATAL	TOTAL	PASSENGER	CREW	GROUND
1970	31	1	9	7	2	0
1971	33	4	201	184	16	1
1972	37	6[a]	65	55	9	1
1973	27	5[b]	139	130	9	0
1974	31	3	168	158	10	0
1975	21	2	122	113	9	0
1976	17	1	1	1	0	0
1977	15	2	75	64	2	9
1978	18	4	164	142	9	13
1979	14	3	278	260	16	2
1980	8	1	13	11	2	0
1981	13	0	0	0	0	0
1982	14	3	233	210	11	12
1983	22	2	2	1	0	1
1984	11	0	0	0	0	0
1985	39	2	172	159	13	0

[a] Includes 05/30 & 12/15 accidents resulting in crew fatalities only

[b] Includes 08/30 accident resulting in one passenger fatality only

TABLE 2 ACTIVITY INDICATORS OF THE U.S. DOMESTIC SCHEDULED AIR CARRIERS

	PASSENGER MILES (18) (billion)	AIRCRAFT MILES (19) (billion)	AIRCRAFT HOURS (19) (million)	NUMBER OF FLIGHTS (20) (million)
1970	104	2.0	6.0	5.0
1971	106	2.0	4.9	4.7
1972	118	2.0	4.9	4.7
1973	126	2.0	5.0	4.8
1974	130	1.9	4.7	4.5
1975	132	1.9	4.7	4.5
1976	145	2.0	4.9	4.7
1977	157	2.1	5.1	4.8
1978	183	2.1	4.7	4.8
1979	209	2.2	5.1	5.0
1980	201	2.5	6.1	5.1
1981	199	2.4	5.9	4.9
1982	210	2.4	5.8	4.7
1983	227	2.5	6.0	4.8
1984	244	2.8^e	6.8^e	5.1^e
1985	270	2.9^e	7.0^e	5.1^e

e estimate

chance. The other is that actuarial data do not take into account the cause of accidents. In the following section, uses, strengths, limitations, and potential misuses of the selected metrics are presented.

AIR TRANSPORTATION RISK METRICS

Fatalities per Hour of Exposure to Air Transportation

The metric numerator *fatalities* is defined by the FAA as injuries that result "in death within 7 days of the accident" (*22*, p. 170). The types of fatalities in air transportation can be broadly classified as on-board fatalities (passengers and crew) and ground fatalities.

The metric denominator *hour of exposure* can acquire different meanings. For example, if the metric is used to measure the safety of the whole flying population, it refers to the number of hours that passengers or crew, or both, fly during a given period of time. If the metric is used to pertain to nonpassengers as well, it refers to the time when there is a danger of an airplane's crashing into people.

The risk metric numerator *fatalities* has some intrinsic strengths and limitations regardless of the choice of denominator. This characteristic will allow a discussion of fatalities as an independent parameter later. This is followed by a discussion of the potential misuses of the entire metric within the framework of specific applications. This approach was chosen because the limitations of the denominator are easiest to recognize when specific uses of the metric are studied.

Uses of the Metric

In 1969 Chauncey Starr published in *Science* his now often-cited article (*23*) Social Benefit versus Technological Risk. In this article Starr used the metric *fatalities per person-hour of exposure* as a wide-spectrum metric to determine the social acceptability of risk. He compared the risks and the benefits derived from hunting, skiing, flying on airliners, flying on general aviation planes, driving, using electric power, and smoking (*3*). *Fatalities per hour of exposure* can also be used as a medium-spectrum metric to compare the safety record of different airlines.

Advantages and Limitations of Fatalities as Numerator

Fatalities as a metric numerator offers some advantages that should not be overlooked. It is an often relevant measure and it is generally reliable in that deaths are always reported. The metric numerator *fatalities* can be easily used in comparisons of the safety performance of different airlines or different modes of transportation.

This metric numerator, however, has a number of intrinsic limitations. Circumstantial factors such as load factors, terrain at the accident site, weather conditions at the time of the accident, and time of day at which the accident happened may impinge dramatically on this measure. The fluctuations of the parameter that may in some cases be attributable to circumstantial factors are shown in Figure 1.

Fatalities as a numerator introduces leveling-of-distinctions and omission bias. Omission occurs because the parameter is not sensitive to severe injuries or to the causes of the different accidents. This lack of sensitivity could have serious consequences, particularly in areas such as the legal determination of compensations and liabilities.

Leveling-of-distinctions bias occurs with respect to life expectancy, willingness to be exposed to risk, airplane size, load factor, and kind of death.

Life Expectancy Before a fatal accident, a 3-year-old passenger has a much longer life expectancy than an 80-year-old; yet *fatalities* overlooks this distinction by simply counting the number of dealths. A legal scholar interested in aviation tort law could not use *fatalities* as a metric numerator because reduction in life expectancy and loss earnings are key elements of wrongful death cases.

Willingness To Be Exposed to Risk *Fatalities* as a metric numerator makes no distinction among the deaths of a crew member, a passenger, and a person on the ground. These three events are quite different in terms of the level of voluntary acceptance of exposure to risk. Crew members constitute the group who subject themselves to the highest risk. Delta Air Lines pilots, for example, fly 75 hr a month and are very much aware of the risks involved in their profession. Passengers are somewhat aware of the risks, and they undertake them on a voluntary basis. Except in those cases in which accidents involve those working at airports, ground fatalities affect those who did not undertake the risk voluntarily.

According to Starr, "the public is willing to accept voluntary risks roughly 1,000 times greater than involuntary risks" (*23*). On the basis of Starr's rule of thumb, one could assume that each ground fatality counted as much as 1,000 passenger fatalities. Under this assumption, from 1970 to 1985 there would have been 41,000 "equivalent ground fatalities," which would surpass the 1,380 actual passenger fatalities (Table 1).

Airplane Size *Fatalities* accounts for a nonsurvivable accident of a fully loaded DC3 in the same way as a survivable accident of a fully loaded L1011 in which 90 percent of the passengers survived. In both cases, the parameter *fatalities* only records the fact that there were 30 casualties.

Load Factors Given two nonsurvivable accidents involving identical airplanes with 100 seats each, one with a 90 percent load factor and the other with a 40 percent load factor, the metric numerator *fatalities* would count 90 fatalities in one case and 40 in the other. The metric numerator *accidents*, however, would regard both events as identical.

Kind of Death The numerator *fatalities* does not distinguish between an instantaneous death and a prolonged agony due to third-degree burns affecting a large percentage of the body. It counts only those deaths that took place within 7 days of a given accident.

Potential Misuses of the Metric

If the metric *fatalities per hour of exposure* were used to compare risks posed by different transportation modes, it would show its leveling-of-distinctions bias by not differentiating between fast and slow modes of transportation. It would discriminate against air transportation by failing to consider the reduction in time of exposure while traveling achieved by the higher speeds of airplanes.

To illustrate how this bias works against air transportation, consider a comparison of the safety levels displayed by National Railroad Passenger Corporation (Amtrak) and Eastern Airlines (EAL) in their Boston–Washington, D.C. (BOS–DC) market. The Amtrak trip lasts approximately 8.5 hr, whereas the EAL trip lasts 1.25 hr. Assuming that both means of transportation produced the same number of fatalities in a given period of time, an individual comparing the safety of Amtrak with that of EAL in the BOS–DC market via the metric *fatalities per hour of exposure* could falsely conclude that Amtrak is 7 times safer than EAL. Using this metric clearly penalizes Eastern Airlines for being a faster means of transportation.

Note that the metric does not consider some important differences between the two modes of transportation in the BOS–DC market, including the number of passenger carried daily, the frequency of departures, and the load factors. Amtrak carries about 8,000 passengers a day, whereas EAL only carries 3,000. Therefore, Amtrak is exposing more than twice the number of people to the risk of traveling seven times longer than on EAL. Amtrak carries an average of 400 passengers per trip, and EAL carries an average of 105 passengers per trip. If it were assumed that the risk was a function of the number of departures, Amtrak would be exposing almost four times as many to the risk of traveling than would EAL. Last, Amtrak offers 20 daily departures in the BOS–DC market, and EAL offers 26. Although the excluded factors are not fundamental in every study, an analyst using the metric *fatalities per hour of exposure* should decide whether those omissions would make a difference in the results.

If one wanted to compare the safety performance of two airlines, one would be better off using another metric, such as a passenger fatalities per 100 million passenger-mi or fatal accidents per flight, because the denominator of the metric *fatalities per hour of exposure* is not optimal in capturing the difference in size and volume of operation of different airlines.

Passenger Fatalities per 100 Million Scheduled Passenger-Mi

If used within aviation, the wide-spectrum metric *passenger fatalities per 100 million scheduled passenger-mi* reflects how many passengers die on average in airplane crashes during the interval in which an airline (or the air transportation system) covers 100 million passenger-mi. The denominator *passenger* *miles* measures travel output as a function of how many passengers fly how far. The numerator of this metric does not consider crew or ground fatalities, which together constituted 9 percent of the total fatalities resulting from the accidents of the U.S. domestic scheduled air carrier operations from 1970 to 1985 (Table 1).

The intrinsic limitations of using *fatalities* as a numerator were discussed earlier. Here the limitations of choosing *passenger miles* as a metric denominator will be examined within the framework of the applications of the risk metric as a whole.

Uses of the Metric

The FAA and the National Safety Council use the metric *passenger fatalities per 100 million scheduled passenger-mi* to compare trends of the normalized accident data of cars, taxis, buses, railroad passenger trains, and domestic scheduled air transport planes (*24*).

The International Civil Aviation Organization (ICAO) compares the safety records of the airlines of the world by using *fatalities per 100 million passenger-km*, which is essentially the same metric.

Potential Misuses of the Metric

A risk analyst must always question the applicability of a particular risk metric for a prescribed use. For example, when comparing airlines, the analyst must be aware that the risk in air transportation is concentrated in the takeoff and landing phases. Data in Table 3 reveal that of the 37 U.S. domestic scheduled air carrier accidents involving passenger fatalities between 1970 and 1985, 27 (73 percent) took place during take-off or landing and 10 (27 percent) took place during cruise. On this basis, it would seem that perhaps a better metric denominator to capture the "exposure" to air transportation risk would be *number of flights* rather than *number of passenger miles.*

In addition, when comparing airlines using *fatalities per 100 million passenger-mi* as a medium-spectrum metric, one has to be aware that the metric favors the larger airlines, which usually have larger airplanes carrying a larger number of passenger for longer distances.

Large airlines complete fewer flights than small airlines to reach 100 million passenger-mi. For example, in 1983 it took United Airlines approximately one day to cover 100 million passenger-mi in its domestic scheduled market. In contrast, it would have taken Mid-South, a middle regional airline, 12.5 years to fly 100 million passenger-mi. During December 1983, United had an average of 1,367 departures a day (or 100 million passenger-mi), whereas Mid-South would have had 119,475 departures in 12.5 years (the required time to perform 100 million passenger-mi). That is, every United takeoff would correspond to 87 Mid-South takeoffs. If one believes that risk in air transportation resides in the takeoff and landing phases rather than in the en-route phase, the last statistic could be interpreted to mean that Mid-South has to undergo a risk 87 times larger than that undergone by United when *100 million passenger-mi* is used as a metric denominator.

Using the metric *fatalities per 100 million scheduled passenger-mi* as a wide-spectrum metric to compare different modes of transportation poses some problems because the differences in speed, passenger-carrying capacity, load factor, and level of service are not accounted for. For example, air transportation is much faster than ground transportation; therefore the time of exposure required to achieve the same number of miles is much less for air transportation than it is for ground transportation. Trains carry roughly 3.5 more passengers per trip than air carriers and 5 times more than a bus. Last, the frequency of departure and availability of service of the different modes of transportation vary considerably depending on the market studied.

Fatal Accidents per 100,000 Flights

The metric *fatal accidents per 100,000 flights* focuses only on those accidents in which human life was lost within 7 days of an air carrier accident. From 1970 to 1985 the U.S. domestic scheduled air carriers completed 100,000 flights in approximately 1 week. This metric represents a good attempt to use a denominator that captures the activity in which the risk resides in air transportation.

Uses of the Metric

Fatal accidents per 100,000 flights is used by the FAA as a wide-spectrum metric to measure the probability of having a fatal accident as a function of the number of takeoffs. The FAA uses this metric in its handbooks of aviation statistics as a safety-trend indicator. It can also be used as a medium-spectrum metric for comparing the safety standards of different airlines.

Advantages and Limitations of Fatal Accidents as Numerator

Fatal accidents is a reliable parameter, because all accidents resulting in loss of human life are reported. One of its limitations, however, is that fatal accidents are rare events. From 1970 to 1985 the U.S. domestic scheduled air carriers experienced 61 fatal accidents. Because of these relatively low numbers, analysts sometimes prefer to use total reported accidents; according to *Flight International*, there were 351 reported accidents for U.S. domestic air carriers from 1970 to 1985.

Fatal accidents as a metric numerator incorporates some leveling-of-distinctions bias because it does not differentiate between a TWA nonsurvivable accident (Dec. 1, 1974) in which 92 lives were lost because of a premature descent (*6*) and a World Airways accident (Sept. 20, 1981) in which there were 345 people on board and only a stewardess trapped in the galley was killed (*14*).

Potential Misuses of the Metric

If cast as fatal accidents per 100,000 departures, *fatal accidents per 100,000 flights* could be used as a wide-spectrum metric to compare different modes of transportation. However, the strong correlation between risk and number of departures is only present in air transportation; other modes show a higher correlation between risk and number of miles traveled.

TABLE 3 FATAL ACCIDENTS OF THE U.S. DOMESTIC SCHEDULED AIR CARRIERS *(2–17)*

		FLIGHT	TOTAL	FATALITIES		
		PHASE	ABOARD	PASSENGER	CREW	TOTAL
01/28/70	TAG	E	9	7	2	9
05/06/71	APACHE	E	12	10	2	12
06/06/71	HUGHES AIRWEST	E	51	44	5	49
06/07/71	ALLEGHENY	L	31	26	2	28
09/04/71	ALASKA	L	111	104	7	111
02/22/72	AIR HAWAII	E	8	7	1	8
06/29/72	NORTHCENTRAL	E	5	2	3	5
06/29/72	AIR WISCONSIN	E	8	6	2	8
12/08/72	UNITED	L	61	40	3	43
07/24/73	OZARK	L	45	37	1	38
07/31/73	DELTA	L	88	83	5	88
08/30/73	TWA	L	141	1	0	1
09/27/73	TEXAS	E	11	8	3	11
11/04/73	NATIONAL	E	128	1	0	1
09/11/74	EASTERN	L	82	69	2	71
12/01/74	TWA	L	92	85	7	92
12/11/74	KEY WEST	E	5	4	1	5
06/24/75	EASTERN	L	124	106	6	112
08/30/75	WIEN ALASKA	L	32	7	3	10
04/05/76	ALASKA	L	50	1	0	1
04/04/77	SOUTHERN	E	85	60	2	62
05/16/77	NEW YORK	L	25	4	0	4
03/01/78	CONTINENTAL	T	197	2	0	2
05/08/78	NATIONAL	L	58	3	0	3
09/25/78	PSA	L	138	129	7	136
12/28/78	UNITED	L	189	8	2	10
02/12/79	ALLEGHENY	T	25	1	1	2
03/10/79	SWIFT AIRE	T	7	1	2	3
05/25/79	AMERICAN	T	271	258	13	271
06/12/80	AIR WISCONSIN	L	15	11	2	13
01/13/82	AIR FLORIDA	T	79	70	4	74
01/23/82	WORLD AIRWAYS	L	208	2	0	2
07/09/82	PAN AM	T	145	138	7	145
01/09/83	REPUBLIC	L	36	1	0	1
10/11/83	AIR ILLINOIS	L	10	7	3	10
08/02/85	DELTA	L	163	128	8	136
09/06/85	MIDWEST	T	36	31	5	36

En route (E); Landing (L); Take-off (T)

Fatal accidents per 100,000 flights could also be used as a medium-spectrum metric to compare two airlines. However, problems may arise if one of the two airlines compared has a significantly larger number of departures or much larger airplanes, or consistently travels with much higher load factors than the other.

Probability of Being Killed in an Air Carrier Accident

The odds of being killed in an airline accident are obviously of interest to the passengers. In this paper, the metric was derived by using statistics on the U.S. domestic scheduled air carriers from 1970 to 1985 (Tables 1 and 2). The fatality quotient was defined as the fraction of passengers who did not survive a given flight; it was set to zero for all the flights that landed safely. The fatality quotient was derived by dividing the total number of passenger and crew fatalities by the total number of passengers on a given flight (Table 4). After all the positive fatality quotients had been calculated, they were summed on a yearly basis. The normalized quotient was derived by dividing the cumulative yearly fatality quotients by the annual number of flights. The normalized quotients reflect the probability that a passenger or crew member will be killed in a U.S. domestic scheduled air carrier accident during a particular year. The odds of being killed in an air carrier accident are shown in Table 4 as the inverse of the normalized quotient.

The metric takes into account the number of survivors of a given accident but not the mileage flown. Figure 4 shows the fatality quotients for each fatal flight of the U.S. domestic (scheduled and unscheduled) air carriers from 1970 to 1985. The fatality quotients take values close to either 0 or 1, which can be easily approximated by a Poisson process. The major limitations of this metric reside in its measurement of risk in terms of the fatality quotients, which depend on the number of fatalities. These limitations were discussed earlier.

Uses of the Metric

The probability of dying in an air carrier accident was used as a risk metric by Barnett et al. (*25*) to compare the safety records of 58 major world airlines.

Potential Misuses of the Metric

Probability of being killed in an air carrier accident could be misused if the limitations of using fatalities as a risk-measuring parameter were not realized. Furthermore, if it were modified to be used as a wide-spectrum metric for comparing the safety of air transportation with that of other modes of transportation, it would lose one of its major attractions, namely, the Poisson process approximation. This would occur because typically a high percentage of passengers survive train and bus accidents, thereby making it inaccurate to assume a 0-1 process. Last, this

TABLE 4 DERIVATION OF PROBABILITY OF BEING KILLED IN AN AIR CARRIER ACCIDENT

	FATALITY QUOTIENT (2–17)	YEARLY FLIGHTS (20) (million)	NORMALIZED QUOTIENT (10^{-7})	ODDS OF BEING KILLED ON A FLIGHT (denominator in millions)
1970	1.00	5.0	2.00	1 in 5.0
1971	3.86	4.7	8.22	1 in 1.2
1972	3.70	4.7	7.88	1 in 1.3
1973	2.86	4.8	5.96	1 in 1.7
1974	1.87	4.5	4.14	1 in 2.4
1975	0.90	4.5	2.01	1 in 5.0
1976	0.02	4.7	0.04	1 in 235
1977	0.89	4.8	1.85	1 in 5.4
1978	1.11	4.8	2.32	1 in 4.3
1979	1.51	5.0	3.02	1 in 3.3
1980	0.00	5.1	0.00	1 in ∞
1981	0.00	4.9	0.00	1 in ∞
1982	1.95	4.7	4.14	1 in 2.4
1983	1.03	4.8	2.14	1 in 4.7
1984	0.00	5.1	0.00	1 in ∞
1985	2.00	5.1	3.92	1 in 2.6

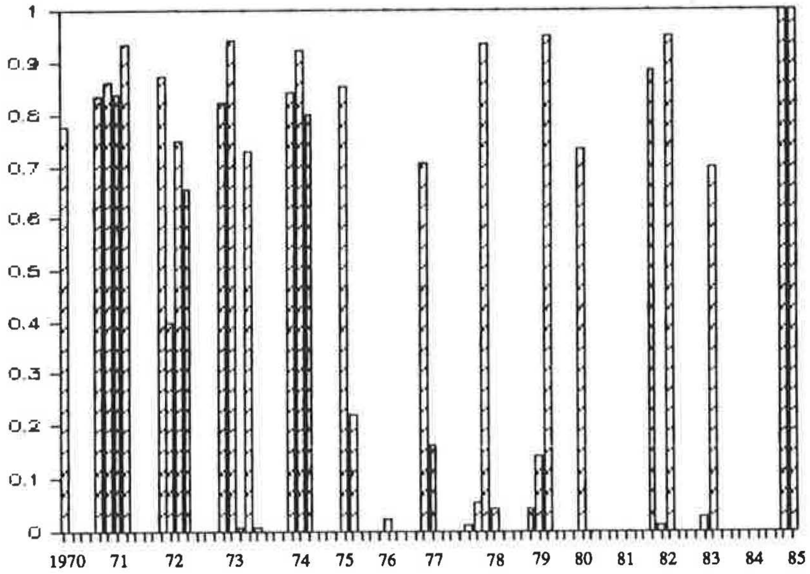

FIGURE 4 Yearly fatality quotients.

metric should not be used if mileage flown is assumed to be a parameter affecting air transportation safety.

Total Accidents per 10 Million System Flying Hours

The metric *total accidents per 10 million system flying hours* records all the air transportation mishaps that resulted in human injuries or damage to airplanes in intervals of 10 million flying hours. During the period studied (1970 to 1985) the U.S. domestic scheduled air carriers flew 10 million hr approximately every 2 years, and there were 351 reported accidents, of which 61 (17 percent) were fatal. This measurement is often used as a medium-spectrum metric for air traffic control purposes. It can be applied to wide-spectrum purposes if it is cast as total accidents per 10 million travel hours.

Uses of the Metric

The North Atlantic System Planning Group (*26*) used the metric in the mid-1960s to set target levels of safety in the North Atlantic region. In its safety trends, the FAA uses a narrower-spectrum metric: *accident rate per 100,000 system hours flown* (*21, 22, 24*). Both these metric denominators (*10 million system flying hours* and *100,000 system flying hours*) encompass a longer time period than 100 million passenger-mi. An attractive feature of *total accidents per 10 million system flying hours* is that it contains a relatively large number of data points.

Advantages and Limitations of Total Accidents as Numerator

Using *total accidents* as a metric numerator represents an effort to include all instances in which something "went wrong" in air transportation. *Total accidents* as a metric numerator incorporates leveling of distinctions because it assigns equal weight to an accident in which an airplane veered off a runway and there were no casualties (Ozark, March 16, 1980) (*28*) and to one in which 273 lives were lost (American Airlines, May 25, 1979) (*11*). Another problem is that the data on total accidents are not completely reliable because not all accidents are reported. Using 10 million system flying hours as a normalizing parameters presents similar problems as using hours of exposure, namely, that faster modes of transportation are discriminated against.

Potential Misuses of the Metric

Total accidents per 10 million system flying hours incorporates serious leveling of distictions by not differentiating between fatal and nonfatal accidents. An analyst interested in loss of life or property should avoid this metric. It should also be avoided if one wants to compare airlines of very different sizes or airlines operating in very different weather conditions.

Miles Flown Between Successive Accidents

To derive the metric *miles flown between successive accidents*, the fatal accidents suffered by both the scheduled and unscheduled U.S. domestic air carriers from 1974 to 1983 were considered. The number of days between successive accidents was counted and then translated to the number of miles flown. This was done by assuming that U.S. domestic air carriers flew the same number of miles every day. The yearly miles flown was divided by 365 (or 366 for leap year) to obtain the daily number of miles flown. The daily mileage was then multiplied by the number of days from one accident to the next (Table 5). When the interaccident time spanned two consecutive years, the number of days corresponding to each year was counted and multiplied respectively by the daily mileage of each of the two consecutive years in question.

Uses of the Metric

This measurement is widely used as a medium-spectrum metric by those concerned with airplane maintenance. It can be cast as *miles traveled between successive accidents* and used as a wide-spectrum metric to compare different modes of transportation. The metric *miles flown* (or traveled) *between successive accidents* has the probabilistic interpretation of "interarrival time" for a stochastic process. On the basis of this notion, a policy maker might want to set safety standards using this metric so that there is a probability not greater than a prescribed target level of having an accident in an interval of a predetermined duration.

TABLE 5 DAYS ELAPSED AND MILES FLOWN BETWEEN ACCIDENTS OF THE U.S. DOMESTIC SCHEDULED AIR CARRIERS

	CARRIER (2–17)	DAYS	MILES (million)
03/13/74	SIERRA		
09/11/74	EASTERN	171	700
12/01/74	TWA	80	425
06/24/75	EASTERN	186	992
04/05/76	ALASKA	302	1638
04/04/77	SOUTHERN	363	2065
05/16/77	NY AIRWAYS	41	243
12/13/77	NATIONAL	221	1308
03/01/78	CONTINENTAL	77	470
05/08/78	NATIONAL	67	413
09/25/78	PSA	139	856
12/28/78	UNITED	93	573
02/12/79	ALLEGHENY	45	303
03/10/79	SWIFT AIRE	25	169
04/18/79	NY AIRLINES	38	257
05/25/79	AMERICAN	36	244
06/12/80	WISCONSIN	383	2611
09/07/81	AMERICAN	451	3058
01/13/82	AIR FLORIDA	127	850
01/23/82	WORLD AIRWAYS	9	60
07/09/82	PANAM	166	1111
08/11/82	PANAM	32	214
01/09/83	REPUBLIC	150	1006
10/11/83	AIR ILLINOIS	274	1915

Potential Misuses of the Metric

The main problem with this metric derives from the differences in severity of aviation accidents. The measured interarrival time may be misleading because it may reflect the time elapsed between events that are very different in nature. An analyst concerned about total levels of air transportation safety may be more interested in total number of fatalities than in the relative frequency of accidents. The media, however, seem to abuse this metric by reporting "new" accidents.

Mean Time Between Failures

This metric is similar to ***miles flown between successive accidents*** and it is the most commonly used narrow-spectrum metric. To derive the mean time between critical failures, analysts must decide which failure to count, for how long, and during which phases. Once analysts have decided which parameter to measure, they must decide how to measure it. Broadly speaking, avionics tests can be divided into destructive and nondestructive ones. Within these two major groups, one can then classify the tests as environmental, physical, and electrical. Once a testing setup is working, analysts have only to count failures during a predetermined period of time and average the interfailure times to obtain the mean time between failures.

A difficult challenge is to understand how the narrow-spectrum metrics relate to wider-spectrum metrics. Ideally, one should be able to translate risk quantitatively among the different sectors of the air transportation industry. Such translations would aid in understanding how narrow-spectrum risks contribute to the aggregate air transportation industry risk.

An example of a probabilistic model that would translate the probability of a critical failure into the expected number of fatalities follows. This model was suggested by B. B. Myers of Transport Canada.

$$E = NP_{F|A}P_{A|CF}P_{CF} \tag{1}$$

where

E = expected number of fatalities,
N = number of occupants (crew and passengers),
$P_{F|A}$ = probability of fatalities given an accident,
$P_{A|CF}$ = probability of an accident given a critical failure, and
P_{CF} = probability of critical failure.

$$P_{CF} = P_s + P_e + P_c \tag{2}$$

where

P_s = probability of a system's contributing to failure,
P_e = probability of the environment's contributing to failure, and
P_c = probability of crew's contributing to failure.

The stochastic quantities involved in Equations 1 and 2 are difficult to estimate. However, any attempt to quantify the probabilities in these equations would prove useful, because it would put things in perspective.

Uses of the Metric

This metric is often used to measure and forecast the reliability of aircraft subsystems, especially that of avionics. According to Bird and Herd (27), avionics accounts for 40 percent of the problems of all aircraft types. *Mean time between failures* is often used in the testing and operational stages of avionics. However, there are limitations in that the reliability of avionics depends on numerous factors, including the complexity and quality of the design, use of state-of-the-art technology, avionics-aircraft interface, resources allocated for avionics development, intensity of monitoring and quality control during manufacturing, training of service personnel, maintenance standards, and test facilities (*28*).

Potential Misuses of the Metric

Mean time between failures might present problems if a careful record of the test conditions and the responses to the test is not kept. Furthermore, this metric should not be used as a deterministic parameter, but rather as a probabilistic parameter with confidence intervals.

COMPARISON OF RISK METRICS EXAMINED

In the previous section the strengths, limitations, uses, and potential misuses of seven commonly used aviation risk metrics were examined. The metrics of interest include fatalities per hour of exposure to air transportation, passenger fatalities per 100 million scheduled passenger-mi, fatal accidents per 100,000 flights, probability of being killed in an air carrier accident, total accidents per 10 million system flying hours, miles flown between successive accidents, and mean time between failures.

Comparing these metrics is difficult because their duration spans are very different. For example, the U.S. domestic scheduled air carriers fly 100 million scheduled passenger-mi in less than 1 day, 100,000 flights in 1 week, and 10 million flying hours in 2 years. One way of avoiding the duration-span incompatibility is to modify the metrics so that they can be expressed as yearly parameters. Figures 5 and 6 show the risk levels displayed by the U.S. domestic scheduled air carriers as measured by the following modified metrics: yearly probability of being killed in an air carrier accident (Parameter A in Figure 5), yearly total fatalities per yearly system flying hours (Parameter B in Figure 5), yearly total accidents per yearly aircraft miles flown (Parameter C in Figure 6), yearly passenger fatalities per yearly passenger miles (Parameter D in Figure 6), and yearly fatal accidents per yearly flights (Parameter E in Figure 6).

One can see that different metrics yield different perceived levels of safety, and the information derived from these tables can be conflicting. For example, from 1972 to 1974, Parameter A in Figure 5 leads one to believe that safety is improving, whereas Parameter B gives the opposite impression. From 1976 to 1985, Parameters C, D, and E in Figure 6 present inconsistent information regarding the relative levels of safety.

To avoid confusion, aviation safety studies should be conducted using several metrics. An understanding of the effects of using different risk metric numerators and denominators is crucial.

CONCLUSION

In this paper an attempt has been made to show that there is no "correct" or "unique" way of measuring risk in air transportation and that the risk metric selected for any study involving risk appraisal may influence the perception of risk. Because of the potential introduction of biases by a particular metric choice, one should always question the applicability and limitations of risk metrics. When possible, aviation risk analyses should be conducted using several risk metrics and should state the applicability, strengths, and limitations of each metric.

One of the weaknesses of this study is that none of the numerators of the risk metrics deal with the underlying cause of the accidents or the fatalities. The metrics make no distinction between accidents caused by weather and accidents caused by pilot error or faulty maintenance. Determining and accounting

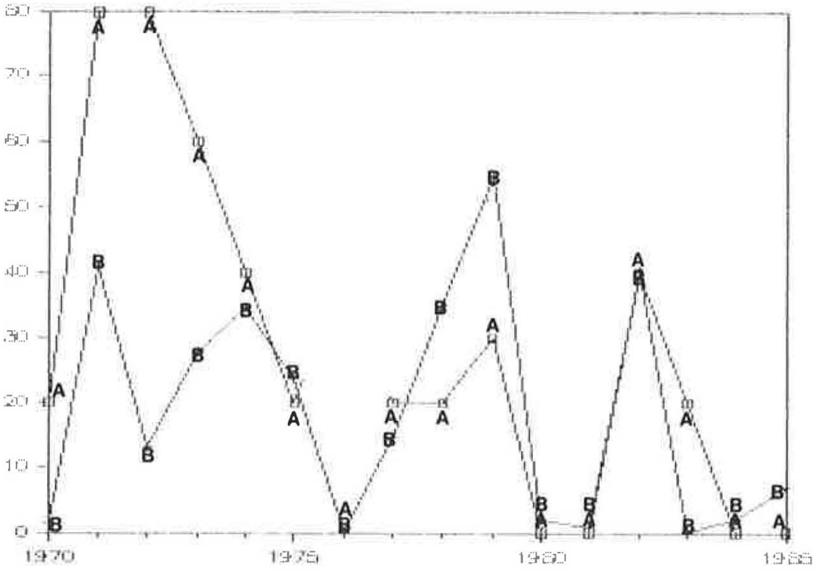

FIGURE 5 Comparison of risk metrics: *A*, yearly probability of being killed in an air carrier accident; *B*, yearly total fatalities per yearly system flying hours.

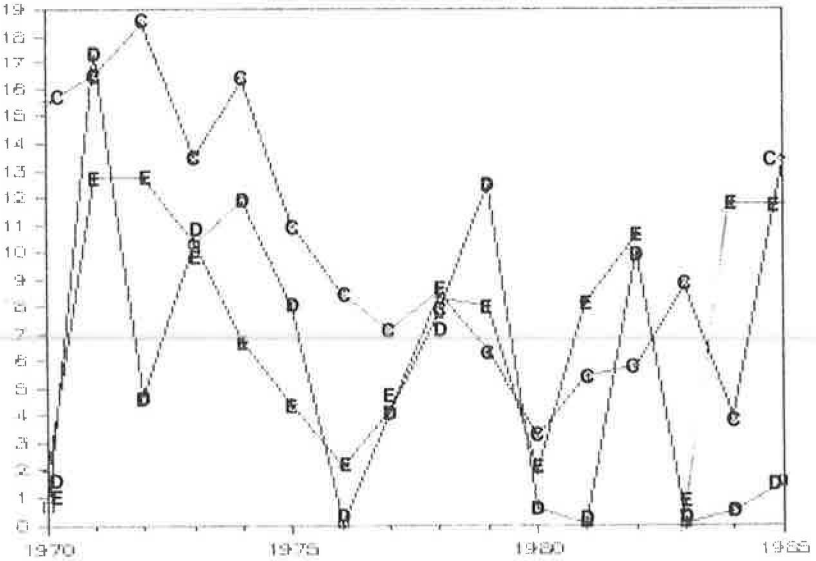

FIGURE 6 Comparison of risk metrics: *C*, yearly total accidents per yearly aircraft miles flown; *D*, yearly passenger fatalities per yearly passenger miles; *E*, yearly fatal accidents per yearly flights.

for the different accident causes is a difficult task. However, it seems desirable to incorporate this notion in order to conduct more equitable and balanced risk analysis studies.

ACKNOWLEDGMENT

This paper was written for the Graduate Research Award Program sponsored by the Federal Aviation Administration and administered by the Transportation Research Board.

The author would like to thank E. Thomas Burnard, Program Officer, Transportation Research Board; Lawrence Kiernan, Airport National Planning Division, Federal Aviation Administration; and Barry B. Myers, Transportation Development Centre, Transport Canada, for their support, encouragement, and help throughout her work in this project.

The research presented in this paper was done at the Flight Transportation Laboratory of the Massachusetts Institute of Technology.

REFERENCES

1. *Federal Aviation Administration Statistical Handbook of Aviation.* U.S. Department of Transportation, 1980, p. 156 (for 1970 to 1979 data).
2. Accident Review: 1970. *Flight International,* Jan. 21, 1971, pp. 105–106.
3. H. Field. Air Transport. *Flight International,* Jan. 20, 1972, pp. 102–103.
4. H. Field. A Small Step Back. *Flight International,* Jan. 18, 1973, pp. 92–93.
5. Safety Record Improves. *Flight International,* Jan. 17, 1974, pp. 81–83.
6. H. Field. The 1974 Fatal Accidents. *Flight International,* Jan. 23, 1975, pp. 115–116.
7. H. Field. Fewer 1975 Fatalities. *Flight International,* Jan. 24, 1974, pp. 182–185.
8. H. Field. 1976 Public-Transport Accidents. *Flight International,* Jan. 22, 1977, pp. 178–179.
9. H. Field and J. Wilkinson. Flight Safety 1977—A Safe Year for Scheduled Passengers. *Flight International,* Jan. 21, 1978, pp. 182–185.
10. H. Field and J. M. Ramsden. Airline Safety 1978: the First Appraisal. *Flight International,* Jan. 20, 1979, pp. 184–187.
11. H. Field. Airline Safety. First Assessment of 1979. *Flight International,* Jan. 26, 1980, pp. 247–253.
12. D. Learmount. Flight Safety: 1980 Reviewed. *Flight International,* Jan. 24, 1981, pp. 233–240.
13. D. Learmount. Commercial Flight Safety: 1981 Reviewed. *Flight International,* Jan. 23, 1982, pp. 183–185.
14. D. Learmount. Airline Flight Safety: 1982 Reviewed. *Flight International,* Jan. 22, 1983, pp. 205–208.
15. D. Learmount. Airline Flight Safety: 1983 Reviewed. *Flight International,* Jan. 28, 1984, pp. 286–288.
16. D. Learmount. Commercial Flight Safety: 1984 Reviewed. *Flight International,* Jan. 26, 1985, pp. 38–42.
17. D. Learmount. 1985: A Turning Point For Safety? *Flight International,* Jan. 25, 1986, pp. 29–30.
18. Passenger Statistics, U.S. Air Carriers Scheduled Service, Domestic and International Operations. *Aerospace Facts and Figures,* Oct. 1986, p. 91.
19. *Air Traffic Statistics.* 1970–1983.
20. *Federal Aviation Administration Statistical Handbook of Aviation.* U.S. Department of Transportation, 1970–1984.
21. *Federal Aviation Administration Statistical Handbook of Aviation.* U.S. Department of Transportation, 1984, p. 24.
22. *Federal Aviation Administration Statistical Handbook of Aviation.* U.S. Department of Transportation, 1983, p. 170.
23. C. Starr. Social Benefit Versus Technological Risk. *Science,* Vol. 165, 1969, pp. 1232–1238.
24. *Federal Aviation Administration Statistical Handbook of Aviation.* U.S. Department of Transportation, 1982, p. 161.
25. A. Barnett, M. Abraham, and V. Schimmel. Airline Safety: Some Empirical Findings. *Management Science,* Vol. 25, No. 11, 1979, pp. 1045–1055.
26. A. C. Busch, R. A. Rigolizzo, and B. F. Colamosca. *Analysis to Support Establishing Safety Criteria for the Study of Separation Standards.* Technical Report. FAA Technical Center, Atlantic City, N.J., 1986, pp. 7–15.
27. G. T. Bird and G. R. Herd. Experienced In-Flight Avionics Malfunctions. In *Avionics Design for Reliability.* Advisory Group for Aerospace Research and Development Lecture Series 81, North Atlantic Treaty Organization, London, 1976, pp. 3.1–3.4.
28. J. E. Green. Reliability Growth Modelling for Avionics. In *Avionics Design for Reliability,* Advisory Group for Aerospace Research and Development Lecture Series 81, North Atlantic Treaty Organization, London, 1976, pp. 2.1–2.8.