Airlines as Baseball Players: Another Approach to Assessing Time-Trends in Aviation Safety

Arnold Barnett, MIT

“Several studies have found that up to 40 percent of people have some degree of anxiety about flying”

***The New York Times*, 7/24/07**

How safe is it to fly?

Well, how should we measure aviation safety?

An e-mail message:

"My name is L.S. I would like to know if you, as an expert in aviation safety, fly regularly."

"You see, I stopped flying about a year ago and this has affected my life in a significant matter. Just one last question: what are the odds of me dying in a plane crash?"

Given that a passenger's greatest fear is of being killed in a plane crash, there is a natural interest in statistics about the likelihood of that outcome.

Measure of Safety Performance
<u>Over a Past Period:</u>

Death Risk Per Randomly Chosen Flight

Question:

If a person chooses a flight at random from among those of interest (e.g. Brazilian jet flights over the period 1990-99), what is the probability that he will not survive it?

Death Risk per Flight:
Scheduled US jet aviation 2000-09

1 in 23 million

(How good is that?)

Well, Barack Obama was elected the 44th president of the US 228 years after the first president was chosen.

That works out to one president elected every 5.3 years (228/43).

Moreover, over 2000-07, there were 4.1 million births per year in the United States.

That works out to 4.1 x 5.3 ≈ 22 million births every 5.3 years in the US , of whom one on average will be elected President.

In other words, an American kid at a US airport today is ***more likely to grow up to be President*** than to perish in an accident on her next jet flight.

(1 in 22 million vs. 1 in 23 million)

Furthermore, at a risk of 1 in 23 million per flight:

Someone who took one flight every day would on average go 63,000 years before dying in an air crash. (Pretty good, eh?)

In the late 1990's, the FAA set a most ambitious safety goal:

Reduce the fatal accident rate for scheduled US aviation by 80% from the 1994-96 level by 2007.

Death risk per Flight, US scheduled aviation, 1994-96

1 in 3.6 million

(*includes both jet and propeller aircraft*)

Death risk per Flight, US scheduled aviation (jet plus propeller), 2007-09

1 in 24 million

Down 85% from 1994-96 level

Mission accomplished!!

However, the *sheer rarity* of fatal accidents on US commercial flights raises the fear that data about them may not reliably illuminate time-trends in the mortality risk of air travel.

For example, suppose that fatal accidents occur randomly over time at an average rate of one per year:

(i.e., on any given day, the chance of a fatal accident is 1/365)

Then, in any particular year, there would be:

- a **37%** chance of **no** fatal accidents
- a **37%** chance of **one** fatal accident
- an **18%** chance of **two** fatal accidents
- an **8%** chance of **three or more** fatal accidents

In short, there would be considerable year-to-year volatility even in the absence of any trends.

It therefore seems useful to pay attention **not just to fatal accidents**, but to lesser untoward events that **caused no deaths** but had the potential to do so.

At the same time, one does not want a "one event, one vote" scheme that **blurs the distinction** between a mildly hazardous event and an extremely dangerous one.

One approach to thinking about this issue arises from a scheme by which the **Oakland Athletics** evaluated the performance of baseball players.

Let me describe it now.

A Definition:

The full-crash equivalent (FCE) of an untoward event aboard a commercial flight is the fraction of passengers who perish because of it.

(If everyone dies, FCE = 1; if 10% die, FCE = 1/10; no deaths, FCE = 0)

In analogy with the baseball scheme, we can calculate the "expected" FCE for any given event based not on what actually happened, but on the average outcome in historical data for very similar events.

Then, we can add up the expected FCE's for *all untoward events* that occurred over a period to get a "luck adjusted" measure for passenger mortality risk.

Example: Category A Runway Incursions

Suppose hypothetically that:

- In recent years, 2% of category A incursions were runway collisions that caused deaths to passengers on scheduled flights

- On average, one of the two planes involved in the collision was a scheduled flight, and the average death toll on that flight was 50% of the passengers

Then the average passenger FCE per Category A incursion would be 50% of 2% = 1% = 1 in 100

(Right?)

(The *actual* FCE of the incursion could be anywhere from 0 to 2.)

Suppose that:

In a certain year, there were seven category -A incursions (none of which caused any deaths).

- **Under the baseball scheme, the seven events would have a total FCE of 7x (1 in 100) = .07.**
- **Thus, they would be treated as equivalent to *one fatal accident that killed 7% of the passengers*.**
- **The fact that there were actually no deaths would be construed as *a lucky outcome* because the seven events all had the potential to cause fatalities.**

In other years, the "baseball-adjusted" FCE score would fall below the actual FCEs in the category A incursions. Those years would be construed as unlucky.

MIT undergraduates Eric Ni, Jason Scott, and Lei Zhu and I considered roughly 5000 incidents and accidents involving US aviation over the years 1994-2008.

We used data from FAA and NTSB.

We used the baseball approach to assess trends over 1994-2008 in air safety for scheduled US passengers.

Doing so required a number of judgment calls, none of which are self-evidently correct.

Two quick examples:

- Runway Collision Risk

- Risk of Midair Collisions

Over 1994- 2007*, there were 199 category-A incursions at US airports, 297 category-B incursions, and 1152 category-C incursions.

There were three fatal runway incursions, one of which killed all passengers on one commercial plane.

Given these data, how should we **allocate** the risk of passenger deaths in runway collisions among different categories of incursions?

- Among category-A incursions only
- Among category A, B, and C incursions, giving **equal weight** to all such events
- Among A, B, and C incursions, giving **greatest weight** to A incursions and **least** to C incursions.

As for midair collisions, there were none involving scheduled US aircraft over 1994-2008.

However:

There were 65 critical near collisions over that period involving scheduled aircraft, and 372 potentially dangerous near-collisions involving such planes *(FAA classifications)*. And there were 37 actual midair collisions involving GA planes.

How should we proceed?

- It seems artificial to estimate midair collision risk as zero because of the perfect actual record over 1994-2009. (Does anyone seriously believe the risk is zero?)

- But introducing a low but positive risk estimate raises very serious issues.

Moving beyond collisions, we had to deal with untoward events of many kinds. Broadly speaking, we asked two questions:

- Did the danger arise from a **natural hazard** outside the aircraft or a **crisis aboard the plane**?

- **Who was primarily at fault?**

  The pilots? Maintenance people? The aircraft manufacturer? Ground personnel? Malicious individuals? **No one?**

Like Goldilocks, we tried to come up with classifications of "similar" events that were neither too narrow nor too broad.

Ultimately, we came up with baseball-adjusted mortality risk estimates for various years.

Some Results for Passengers on Scheduled US Flights:

Death Risk per Flight:

Year	Actual	Baseball-Adjusted
1995	**1 in 6.6 m**	**1 in 6.3 m**
1999	**1 in 129 m**	**1 in 5.8 m**
2003	**1 in 9.3 m**	**1 in 10.2 m**
2007	**0**	**1 in 14.8 m**

In considering this approach, one does well to remember President Obama's warning that:

The perfect is the enemy of the good.