

# *Airline Health and Safety*



**Arnold Barnett**  
**David Czerwinski**  
**MIT**

*FAA has asked us to assess “the health of an airline from the perspective of safety.”*

*Are there statistical indicators akin to cholesterol scores that suggest whether an airline is “headed for trouble?”*

**In assessing an airline's safety status, there are two distinct needs:**

***Input measures***, which reflect the caliber of its attempts to maintain safety

***Output measures***, which reflect how well it actually succeeds in avoiding adverse events

We turn first to  
*output measures.*

# Question:

Among apparently-similar  
US airlines, are some  
*systematically safer than*  
*others* over the long-run?

A natural place to focus first is on *fatal accidents*. But the rarity of these events, though highly desirable, creates a *great deal of statistical “noise”* in the data about them.

In total, these seven carriers suffered **6.40** full-crash equivalents (**FCE's**) caused by accidents over the 20-year period, out of **83 million flights**.

(In computing FCE's, we weight each plane crash by the **proportion** of those on board who perished in the accident.)

**This is an excellent overall record, implying a passenger death risk per flight of 1 in 13 million.**

**But how were these 6.4  
full-crash equivalents  
*distributed across the seven  
airlines?***

*Accident-caused Domestic Full-crash  
Equivalents for the seven carriers, 1983-2002*

<u>Airline</u>	<u>Proportionate Share</u>	<u>Actual FCE's</u>
AA	1.05	0.08
CO	0.58	0.34
DL	1.24	0.97
NW	0.70	1.17
UA	1.01	1.39
US	1.02	2.46
WN	0.82	0

Among the seven airlines, *actual* FCE's ranged from *0%* (*WN*) to *241%* (*US*) of expected FCE's based on proportionate share of exposure. What can we say about this dispersion?

Well, these variations between proportionate shares and actual FCE's could *easily be explained as chance fluctuations* associated with rare events.

But the fact that they *could be* explained as fluctuations does not mean that they automatically *should be*.

**For another perspective on the “equal safety” question, we used a technique for evaluating baseball players that was described in the book Moneyball.**

## One Ultimate Goal:

A *family of output indicators* that are individually imperfect but that, *taken together*, provide useful information about safety performance on an air carrier, both in absolute and relative terms.

Might the *Educational Testing Service* offer a useful idea about FAA oversight procedures?

Could we achieve such a  
“adaptive” test regimen in the  
context of *aviation safety input  
processes*?

- How might the idea work in (say) assessing the caliber of an airline's *maintenance manuals*?
- How would the idea work in assessing the power of an airline's *data mining system*, which combs reports about recent events for patterns that might require remedies?

*Might the results on these tests--like a student's score on the GRE--in themselves constitute valuable performance measures?*

This “adaptive test” format might not work well in actual practice. But why not *try some experiments* to see what can and cannot be achieved?

*(We’re about to start.)*