Failure Interval Probabilistic Analysis for Risk-based Decisions – Concorde Crash Example

Jan B. Smith, PE

Key Words: Poisson probability, simulation, repairable system reliability, time between failure, system safety, event interval, probabilistic analysis, probability map

SUMMARY & CONCLUSIONS

The DC 6, DC 8, DC 10, Concorde, Boeing 787, and Boeing 737 MAX fatal crashes and near misses were analyzed with event interval probabilistic analysis methods. Fleet grounding decisions are the epitome of risk-based decisions and the most important decision is the first opportunity to ground. The first opportunity to ground decision is retrospectively judged to be wrong if in the immediate future another accident or cause and effect findings leads to the original decision being reversed. Using only data that was available at the time of the significant events, the analysis examines these risk-based decisions as if it they were made at the event instant in time. The event interval method identified five out of six first opportunity to ground decisions correctly, including the Concorde. The FAA and its predecessor organizations made one correct decision out of five. Use of the method based on statistics and probability would have avoided 503 actual fatalities plus 9.45 expected value fatalities from additional risk exposure due to flying statistically proven unreliable aircraft. In addition to the reversed decision standard for judging whether decisions are wrong, a grounding of the DC 8 and a second grounding of the DC 6 are shown to be have been statistically appropriate, but these groundings did not occur.

A specific objective of this paper is to lead the FAA and aircraft manufacturers into using event interval probabilistic analysis in grounding decisions and air worthiness certification. Cause and effect data that are necessary to identify issues and make corrections are sparse or nonexistent at the time of the event. Cause and effect data can take days or months to acquire and analyze, but event interval timing data are simple, system performance data and are available at the instant the event occurs.

The method is appropriate to all engineered repairable systems and has been applied to some non-engineered systems. The various applications are left to the creativity of the reader. Here we focus principally on commercial aircraft. An aircraft type is considered a system composed of, for example, design, manufacturing, documentation, training, and field support. Anything that touches on reliability is within the system boundary.

Statistics and probability analysis demand much data; however, for serious critical events failure and accident data must be few. This conflict is resolved by a null hypothesis that the data are generated by a homogeneous Poisson process (HPP). The analysis uses the infinite quantity of data inherent in this null hypothesis. Event data are compared with the null hypothesis and the null is rejected or not with Poisson and/or computer simulation probability values (p-values). The Poisson interval is not limited to time. For aircraft accidents, the number of departures between events (DBE) are used. Departure counts are for the aircraft type and/or individual aircraft. Untimely events that denote statistically significant deviation from expectation are distinguished from random variation in an expected system. In this paper, expectation is the worldwide fleet fatal accident rate for commercial jet service, except where noted. This paper demonstrates the analysis method and value by analysis of the Concorde crash, selected to demonstrate the analysis power with only one event. This example also demonstrates that the high-level system raw data required for analysis are always available for any system important enough to be of interest.

Decisions made with p-values involves the risk of false positives. The sum of all p-values that suggest grounding is appropriate for all the fleets examined over the 75-year period sum to only 4.67% chance of a single unnecessary grounding. This is small in comparison to, say, not grounding the 737 MAX after the first crash. The risk of British Airways continuing to fly for an additional 21 days before grounding the fleet is analyzed in detail.

All aircraft analyzed are shown to have been placed into service while unreliable as measured against the worldwide commercial fleet (domestic fleet for the DC 6). Great strides have been made to reduce this worldwide fatal accident rate, but for the systems analyzed, some of the improvement is a result of lessons learned from serious, often fatal, accidents. Of course, we learn from accidents, but better yet not to have the accident. The analysis method can help keep low reliability systems from entering service. When certifying air worthiness, the FAA is encouraged to set p-values below which fleets will be automatically grounded, in the absence of strong cause and effect data to the contrary. This performance requirement will serve as incentive to ensure that total system reliability is considered, including pilot error, human factors, and training, for example. A net benefit will accrue to the aircraft manufactures as the low false positive rate is trivial compared to the impact of low reliability aircraft being in service.

1 *INTRODUCTION*

Failure and accident events on important systems often are very few, e.g. 0, 1, 2, …, while data analysis with statistics and probability typically require a significant quantity of data. Event interval analysis, or failure time analysis as it is often applied, is conventionally applied to datasets and is not applicable to very small datasets. This conventional method is applicable to larger datasets as the DC 6 and DC 8 analysis will illustrate. The nonconventional event interval probability analysis method is applied to all aircraft types and events. It is applied to individual events and contiguous groups of events, such as the Boeing 737 MAX first event, second event, first and second events combined, and risk of a third event. This is to identify an event interval or contiguous group of event intervals that signal a statistically significant deviation from the expected or target rate.

A homogeneous Poisson process (HPP) is taken as a null hypothesis with the null having a mean interval equal to the expected. The expected interval for aircraft used in this paper is mean departures between fatal events (MDBE) for the worldwide commercial jet fleet at the time of the events under analysis. The alternative hypothesis is the event interval(s) is less than that of the null. The null is rejected on the strength of probability values (p-value) from either Poisson probability or computer simulation.

Poisson and simulation provide nearly identical p-values, although these are totally different concepts. While different in concept, they share common underlying mathematics. Both methods are required. Each accomplishes tasks not possible or practical by the other. Application to the Concorde event will demonstrate method. Input data, results, and grounding decisions are shown for all systems and events analyzed.

Probability maps are presented to organize the numerous related equations and present results in an understandable manner, illustrated with a process pump dataset example.

2 *PROBABILITY BY SIMULATION*

The null hypothesis against which we test event interval data is an HPP. For the null, we say the system under analysis generates events so that there is a constant event rate over time. The average number of events over a fixed time interval is unchanged over time. All variation in event intervals from one to another in this null hypothesis system is 100% due to only natural randomness with 0% due to real system changes. Real data from real systems are never a perfect HPP as cause and effect impacts would have to be perfectly frozen. This is impossible, so perfect HPP event data must always be computer generated. Our null hypothesis is simply a convenient mathematical abstraction. It provides an infinite quantity of perfect data that has a perfect absence of cause and effect influences on event intervals. Real events are compared with the null generated events. If the real event "fits in" with the null generated events, we do not reject the null. If the real event seems to be an outlier, it does not "fit in" with the null generated population of event intervals. The degree to which real events "fit in" is quantified with p-values. These measure the probability that the real event intervals are not consistent with natural randomness in the null hypothesis. From a quality control or statistical process control viewpoint, this is a way to distinguish special cause events from common cause events. The Boeing 737 MAX data have been plotted on an SPC control chart. The event intervals data fall into the rejection regions of the control chart [12].

Figure 1 describes the basic process for Monte Carlo simulation. The cumulative failure distribution, equation 1, is the complement of reliability, or unreliability. A uniformly distributed zero to 1 random number is transformed to a time sample, or DBE sample, by equation 2 and 3, respectively. Figure 1 illustrates the basic process.

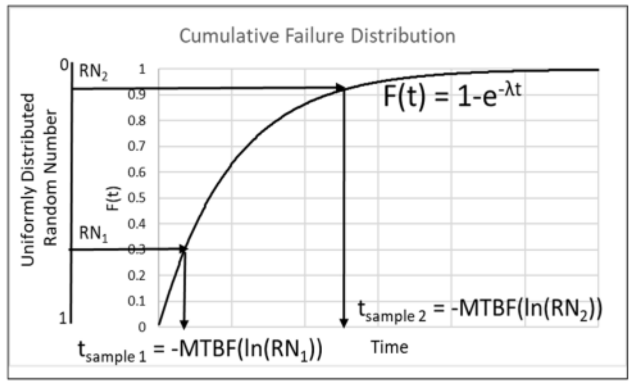

Figure 1
A failure time sample is determined with a random number draw that is transformed to a failure time by the failure distribution. The graph shows two examples of failure times sampling.

$$F(t) = 1-e^{-\lambda t} \tag{1}$$

$$F(t) = 1-e^{-\lambda t} = RN$$

RN = uniformly distributed random number from 0 to 1

$$e^{-\lambda t} = 1-RN$$

the complement of a random number is a random number

$$e^{-\lambda t} = RN$$

$$-\lambda t = \ln(RN)$$

$$\lambda = 1/MTBF$$

$$t = -MTBF*(\ln(RN)) \tag{2}$$

changing time to departures between events (DBE) and MTBF to MDBE,

$$DBE = -MDBE*(\ln(RN)) \tag{3}$$

The Concorde fleet consisted of a small number of planes operated by Air France and British Airways each operated seven airplanes. At the time of the Concorde crash, the Concorde fleet had accumulated 83,941 departures[8].

The commercial worldwide jet fleet mean departures between fatal events at the time of the Concorde crash was 4,000,000[4], as approximated from a chart. This is the value of MDBE in equation 3. Numerous DBE samples make up the distribution of figure 2. The number of departures for the

Concorde system at the time of the crash is marked in figure 2. The fraction of the area to the left of the marker is the probability of failure by 83,941 departures and is 0.02124. (Throughout this paper, estimates and calculated values are shown to several decimal places. This is to help those duplicating the calculations to confirm method and does not imply significance at, say, five decimal places). This probability value is sufficiently low to reject the null hypothesis. There is about a 1 in 50 chance that a value as low as 83,941 would be experience by simply random variation of an otherwise reliable system (as reliable as the null hypothesis). With the null rejected the alternative hypothesis is accepted. The system is statistically less reliable than the worldwide fleet. We act and make decisions as though it is an unreliable system. We cease to view the event as unfortunate randomness within a reliable system. Randomness is distinguished from true deviation from the standard with statistics and probability. There is always a chance of being wrong with this decision. In this case, the probability of rejecting the null when it is true, i.e., a false positive, is 0.02124.

Air France and British Airways both had seven aircraft in service. A crash happened on July 25, 2000 and the Air France fleet was grounded immediately. British Airways continued to operate the fleet for 21 days before grounding. The risk of a second crash within the 21 days is significant. The alternative hypothesis is that the fleet is significantly less reliable than the standard. To measure the risk associated with the 21 days of operation with low reliability aircraft, a distribution of DBEs for the low reliability aircraft type is needed.

Equation 3 provides DBE samples from a population with known mean. The mean is a parameter. With the null rejected, the mean is unknown. The DBE of 83,941 is a sample from an unknown mean. It is a statistic. The unknown mean is a random variable – a distribution. If there are numerous samples, the distribution will be narrow, otherwise the distribution is widely spread. Equation 3 is used with the mean 83,941 to generate sample DBEs. If, for example, if there are five events, equation 3 is used to generate five consecutive event intervals that are averaged. This number is used as a sample of the unknown MDBE. A numerous collection of these averages provides a distribution of the statistic MDBE as a random variable, or $MDBE_{RV}$. Equation 3 is used to produce $MDBE_{RV}$ in equation 4. In the special case of only one event, the distribution for $MDBE_{RV}$ is identically equal to the distribution of figure 2, because an average of a single number is the single number. Equation 4 is equation 3 used twice within the same calculation, with two different and independent random numbers.

$$DBE = -MDBE_{RV}*(\ln(RN)) \qquad (4)$$

In summary, equation 3 with our statistic mean is used to generate the number of DBE values equal to the number of events in the data. These events are averaged and are one sample of the $MDBE_{RV}$ in equation 4. Then equation 4 provides a sample DBE for the failure distribution, that is figure 3 for the Concorde. The null hypothesis, figure 2, is rejected by a low p-value, and the alternative hypothesis, figure 3, is accepted. This is accomplished with a dataset that, in this case, contains only the one failure - thanks to a null hypothesis that provides an infinite quantity of perfect data.

Air France grounded their fleet immediately upon the crash, but British Airways operated their fleet for an additional 21 days. With an assumed one flight per day for seven aircraft, the 147 departures applied to the figure 3 data gives a probability of a second crash of 0.0112. Of course, this high risk was unmeasured and not recognized, as has always been the case so far. In figure 3, notice the wide distribution with a long left-tail, even with a log scale that visually minimizes risk at the lower departure values. Additional failure data will tighten the distribution; however, in service failure of critical systems do not allow this. We must develop and use methods applying the failure and accident data that we unfortunately have with the goal of reducing and eliminating even this minimal data.

Now we consider the risk of a second Concorde fatal accident within the additional 147 departures. The consequence

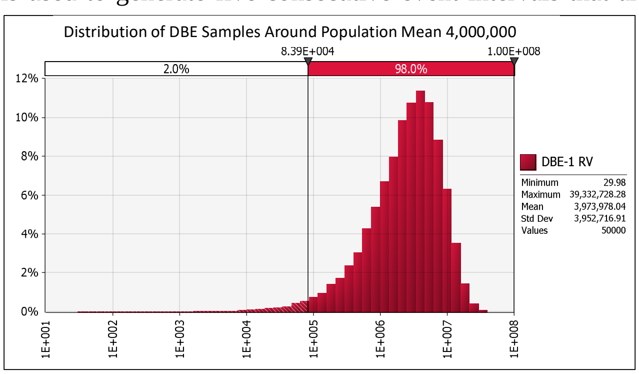

Figure 2
Worldwide commercial jet fleet DBE distribution around the mean of 4 million departures. This is the null hypothesis against which event data are compared. The probability of an event by 83,941 departures is 0.02124. This low p-value should reject the null. The crash cannot be treated as an unfortunate accident of an otherwise reliable aircraft.

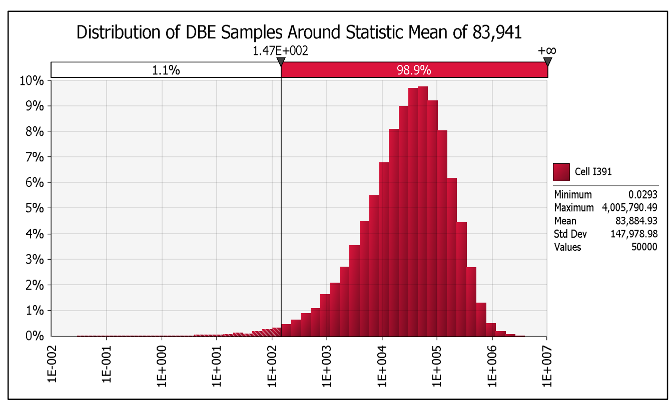

Figure 3
With the null rejected, the failure distribution for the now proven unreliable aircraft is determined from equation 4. The probability of a second crash by 147 departures is 0.0112, the fraction of area to the left of the marker. The long left-hand tail indicates early low-level probability of an additional event.

of a second event can be said to be either 0 or 109 fatalities, the number in the first crash. Actual experience will be one or the other, but neither reflects true risk. Risk is the product of probability and consequence; therefore, risk is 0.0112 probability from immediately above multiplied by 109 fatalities or 1.22 virtual fatalities. This risk would never be taken for the benefit of 21 days of operation if it had been recognized. Event interval probabilistic analysis will permit better risk-based decisions by exposing this risk.

3 POISSON PROBABILITY

Poisson probability is a different event interval concept that confirms the simulation method and allows practical implementation not feasible with simulation. The Poisson distribution is used to determine the probability of specific numbers of events occurring within a specified time interval, when the events are generated by an HPP. Failures and time are common Poisson events and intervals, respectively. But for commercial aircraft, departures are the better event interval measure. The airplane systems and events covered span 75 years so various types of data are available. Departures were used when available or could be estimated; otherwise, various ways are used to get system intervals consistent with the population mean unit of measure. The source of raw data and basis for estimates are noted in footnotes in Table 1. The Poisson probability distribution of events is:

(1)

Where:

e: An approximately 2.71828 constant, the base of the natural logarithm system.

μ: The mean number of events expected in an interval

x: A specific number of events in the interval

$P(x;\ \mu)$: The probability that x events are experienced, given the mean number expected is μ.

The general Poisson expression is now adapted specifically to failure events.

$$\mu = t/MTBF \qquad (2)$$

Where:

MTBF = mean time between failure, the reciprocal of failure rate.

t = time period, a time between failure (TBF) value of interest or sum of one or more consecutive TBF values.

The Poisson distribution for failure events gives the probability of any specific number of failures x and is dependent on the time interval and MTBF, as below:

$$P(x;\ t/MTBF) = (e^{-t/MTBF})(t/MTBF)^x/x! \qquad (3)$$

Changing common nomenclature to that appropriate for aircraft:

$$P(x;\ DBE/MDBE) = (e^{-DBE/MDBE})(DBE/MDBE)^x/x! \qquad (4)$$

For the following Poisson probability results, p-v1 is the probability value for one or more events occurring in the DBE Poisson interval for that one event, as described in earlier papers[1,2]. The probability of one or more events within the interval experienced is the complement of the probability of zero events in the interval, therefore:

$$\text{Concorde crash: p-v1} = P(x \geq 1) = 1 - P(x=0) = 0.02077$$

This is essentially the same as the simulated value. Just as the very low simulation p-value shows the fleet should be grounded, the Poisson concept renders the same decision with low p-values.

3 AIRCRAFT FLEET GROUNDING DECISIONS

Six grounding decisions by the FAA or their predecessor organization, CAA and DGAC (UK and France equivalent, respectively) were revisited with both the computer simulation and Poisson probability methods. These decisions and the analyses used in the decision review are obviously applicable to any engineered system. The surveyed decisions covered 75 years, five aircraft types and 30 p-values with all but four rejecting the null hypothesis and calling for fleet grounding. P-values would ground immediately upon 9 of the 10 events. The only event that p-values would fail to ground when grounding was the correct decision was a DC 6 crash in 1947. The first opportunity to ground is the most important decision. The FAA made the correct decision to ground upon the DC 10 pylon failure in 1979. The other four first opportunity to ground decisions were proven wrong by events shortly after the decision. Event interval probability analysis suggested the correct decision to ground in five out of six cases. A p-value of 0.237 failed to reject the null hypothesis upon the first crash of the DC 6.

The above decisions are judged wrong only if they are reversed due to future events. Groundings that perhaps should have occurred but did not – a second DC 6 certainly and possibly a DC 8 grounding - are not included in the above and are reviewed later.

Input data and results are shown in table 1. Reading table 1 requires defining the terms “p-value1 and p-value2”. P-value1, usually shortened to p-v1, for any event is the probability of one or more events occurring in the Poisson interval of that event. P-value2, or p-v2, is the probability of two or more events occurring within the Poisson interval of the event at which p-v2 is assessed plus the prior event. Probability maps are introduced later. These have equations that will assist understanding.

Of course, the use of p-values in grounding decisions means there is a chance of a false positive that will unnecessarily ground a fleet. The sum of all the first opportunity to ground p-values in table 1 is 0.0467. The 4.7%

Table 1

Summary of data sources and estimates, p-values, grounding decisions and consequences. P-values in red font are those that should reject the HPP null hypothesis, i.e., departures between events indicate statistically significant unreliability relative to contemporary aircraft.

Plane[1] / Event[2]	DBE[3]	MDBE[4]	Poisson p-value1	Poisson p-value2	Simulation p-value1	Simulation p-value2	Grounding by FAA, CAA or DGAC	Grounding by Event Interval Analysis	Actual Fatalities	Probability of an Additonal Event	Avoidable Fatalities - Actual & Virtual
Boeing 737 MAX											
(10/29/18) 1st Event	135,980[5]	6,105,714[6]	0.02202		0.02251		No	Yes	189		
(3/10/19) 2nd Event	139,313[6]	6,105,714	0.02256	0.00099	0.02290	0.00077	No	Yes	157		157
(plus 3 days) Virtual 3rd Event	3,549[6]	137,647[10]					Yes (late)	Yes		0.0475	8.22
Boeing 787 battery fire											
(1/7/13) 1st Event	12,320[11]	5,000,000	0.00246		0.00249		No	Yes	0		
(1/15/13) 2nd Even	704[13]	5,000,000	0.00014	0.000003	0.00015	0.00001	Yes	Yes	0		
Concorde											
(DGAC France) 1st Event	83941[17]	4,000,000	0.02077		0.02124		Yes	Yes	109		
(CAA UK) 1st Event	83,941	4,000,000					No				
(CAA plus 21 days) Virtual 2nd Event	147[16]	83,941					Yes (late)	Yes		0.0112	1.22
DC 10 Door[18]											
(6/12/72) 1st Event	8504[7,19]	500,000	0.01686		0.01626		No	Yes	0		346
Specific aircraft (6/12/72) 1st Event	638[7,19]	500,000	0.00128		0.00135						
(3/3/74) 2nd Event	106832[7,19]	500,000	0.19238	0.02285	0.18860	0.02244	Yes	Yes	346		
DC 10 Pylon											
Specific aircraft (5/25/79) 1st Event	110[9,19]	666,666[8]	0.00016		0.000160		Yes	Yes	273		
DC 6 - Fire[18]											
(10/24/47) 1st Event	9,150[14]	33,854[15]	0.23683		0.23418		No	No	52		
(11/11/47) 2nd Event	1264[19]	33,854	0.03665	0.03864	0.03590	0.03900	Yes	Yes	0		

footnotes: (Full details on sources and estimates at www.pmfseries.com)

1- Aircraft type associated with the event(s) leading to fleet grounding decision.
2- The crash or other event related to a grounding decision.
3- Estimated Departures Between Events for the fleet at the time of the event.
4- Mean departures between events for the worldwide jet commercial fleet at the time of the event, from ref. 4 page 16 graph, unless otherwise indicated.
5- Wall Street Journal article link reference 11.
6- From reference 2.
7- Departures to first event for oldest plane in fleet with wearout failure mode.
8- Read from Reference 4 page 16 graph - 1.5 per million in 1979
9- From ref. 11, new procedure introduced on accident aircraft 55 days earlier time estimated 2 flights per day.
10- With null rejected, statistic MDBE is mean of first two events.
11- Fleet days from delivery to 1st event multiplied by an estimated 2 departures per day per plane.
12- Delivery mo. and yr. from planespotters.net with day 15 (mid month) used in calcs.
13- 44 planes * 8 days since first event * estimated 2 departures per day per plane.
14- Reference 9 page 62, DC6 plane-days in service determined with same method as for 737 MAX in ref. 2.
15- Reference 9 pages 62 and 92, 742 planes * 365 / 8 fatal accidents
16- 21 days multiplied by 7 planes in British Airways fleet with estimated 1 flight per day.
17 -Reference 6 page 146.
18- Reference 6.
19- details on estimates can be found at www.pmfseries.com

chance of a single error over a span of several decades is small compared to the 503 actual fatalities and 9.45 virtual (expected value) fatalities that were avoidable. Also, this is an extremely small risk relative to the financial impact of the 737 MAX not being grounded upon the first crash and the unrecognized high risk of a third event that exposed Boeing and the FAA to exponentially even more severe impacts.

4 PROBABILITY MAPS

Probability maps[1, 2] are a computer spreadsheet for the organization of input data and numerous interrelated equations to produce and display event interval probabilistic results efficiently and effectively. The simple and straight forward calculations, as in the Concorde example above, are impractical with larger data sets without a probability map or some equivalent method not yet developed.

A pump failure dataset from the process industry provides a probability map in the most general-purpose form, figure 4. (The DC 6 and DC 8 probability maps in special purpose form are reviewed later). There is no reasonable population to serve as the null hypothesis; therefore, the pump MTBF at the time of each failure will serve as the null hypothesis. This MTBF is a sample statistic. The true pump MTBF is unknown. Most engineered systems will not have a population of contemporaries to provide a single-valued null hypothesis as with commercial aircraft. Most will have MTBF as a random variable. Even when a contract requirement or project goal can serve as a null, there will be no similar null for precursor events, so figure 4 will be frequently used. A characteristic equation is shown for each function. These are in Microsoft Excel form to leverage their efficiency. Probability maps can be formed by understanding the map, then copying down and across making any minor corrections. (The probability map showing all equations can be found at www.pmfseries.com).

As both the Concorde and the pump example illustrate, the event interval methodology described provides decision-making information immediately upon a single event or small

Column/ Row	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
3	Failure	TBF	MTBF	Cum TBF	LaPlace p-value	p-v1	p-v2	p-v3	p-v4	p-v5	p-v6	p-v7	p-v8	p-v9	p-v10	p-v11	p-v12	p-v13
4	1	13	13.00	13		0.6321												
5	2	99	56.00	112		0.8293	0.5940											
6	3	885	332.33	997	0.9839	0.9303	0.7949	0.5768										
7	4	759	439.00	1756	0.9575	0.8225	0.8878	0.7575	0.5665									
8	5	60	363.20	1816	0.7640	0.1523	0.6586	0.8469	0.7299	0.5595								
9	6	503	386.50	2319	0.7695	0.7279	0.4276	0.6642	0.8210	0.7104	0.5543							
10	7	761	440.00	3080	0.8467	0.8226	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]						
11	8	1308	548.50	4388	0.9419	0.9079	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]					
12	9	259	516.33	4647	0.8605	0.3944	0.8060	0.8274	0.7964	0.6577	0.7080	0.7727	0.6731	0.5443				
13	10	8	465.50	4655	0.6741	0.0170	0.1133	0.6571	0.7375	0.7280	0.5902	0.6690	0.7573	0.6640	0.5421			
14	11	38	426.64	4693	0.4704	0.0852	0.0054	0.0360	0.5226	0.6524	0.6653	0.5329	0.6352	0.7439	0.6562	0.5401		
15	12	46	394.92	4739	0.2975	0.1100	0.0197	0.0018	0.0129	0.4103	0.5746	0.6083	0.4832	0.6051	0.7320	0.6493	0.5384	
16	13	11	365.38	4750	0.1604	0.0297	0.0110	0.0024	0.0002	0.0035	0.3092	0.4975	0.5512	0.4345	0.5755	0.7211	0.6432	0.5369
104																		
105	Failure	TBF	MTBF		RV MTBF	TBF-1	TBF-2	TBF-3	TBF-4	TBF-5	TBF-6	TBF-7	TBF-8	TBF-9	TBF-10	TBF-11	TBF-12	TBF-13
106	1	13	13.00		8.5	8.5												
107	2	99	56.00		85.6	171.0	0.1											
108	3	885	332.33		319.4	87.8	170.4	700.0										
109	4	759	439.00		553.0	39.6	688.9	643.6	839.9									
110	5	60	363.20		160.1	87.4	369.2	207.2	114.3	22.1								
111	6	503	386.50		458.5	71.4	48.2	262.9	454.4	1755.1	158.9							
112	7	761	440.00		806.1	1239.8	361.4	62.9	2179.2	439.1	1253.3	107.2						
113	8	1308	548.50		476.6	1012.5	475.8	302.2	158.0	893.9	251.1	707.3	12.3					
114	9	259	516.33		575.2	1308.0	589.0	245.1	149.0	268.1	515.2	242.9	721.2	1138.1				
115	10	8	465.50		913.5	194.4	750.4	326.4	587.8	443.0	1623.3	1868.8	1379.3	1199.7	761.5			
116	11	38	426.64		465.9	63.9	119.9	196.3	93.2	457.8	127.0	1327.1	352.3	929.5	1168.2	289.3		
117	12	46	394.92		280.8	167.5	9.4	497.3	535.9	384.1	596.5	138.8	240.0	427.1	47.4	3.6	321.7	
118	13	11	365.38		357.0	1.9	36.1	132.3	26.3	1.3	1407.0	849.1	271.7	762.1	0.7	194.6	141.9	815.9
206																		
207	Failure	TBF	MTBF		MTBF RV	p-v1	p-v2	p-v3	p-v4	p-v5	p-v6	p-v7	p-v8	p-v9	p-v10	p-v11	p-v12	p-v13
208	1	13	13.00		8.5431	0.7817												
209	2	99	56.00		85.5825	0.6855	0.3762											
210	3	885	332.33		319.3745	0.9374	0.8126	0.6035										
211	4	759	439.00		553.0216	0.7465	0.7967	0.6099	0.3920									
212	5	60	363.20		160.0502	0.3126	0.9633	0.9984	0.9960	0.9881								
213	6	503	386.50		458.4800	0.6662	0.3475	0.5502	0.7078	0.5647	0.3942							
214	7	761	440.00		806.1224	0.6109	0.4647	0.2277	0.2605	0.3093	0.1851	0.0928						
215	8	1308	548.50		476.6406	0.9357	0.9304	0.9050	0.8008	0.8372	0.8825	0.8090	0.6997					
216	9	259	516.33		575.1753	0.3626	0.7557	0.7688	0.7239	0.5641	0.6082	0.6723	0.5549	0.4185				
217	10	8	465.50		913.4572	0.0087	0.0352	0.2492	0[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
218	11	38	426.64		465.8641	0.0783	0.0046	0.0289	0.4552	0.5762	0.5821	0.4425	0.5377	0.6481	0.5478	0.4262		
219	12	46	394.92		280.7890	0.1511	0.0367	0.0046	0.0383	0.7025	0.8591	0.8937	0.8308	0.9143	0.9659	0.9469	0.9109	
220	13	11	365.38		356.9911	0.0303	0.0115	0.0026	0.6790	0.0039	0.3277	0.5216	0.5771	0.4613	0.6044	0.7476	0.6736	0.5701

Figure 4

Pump probability map. Time between failure (TBF) for each failure are entered starting at cell C4. Each line of data considers only that line and history to that point in time. At failure number 10 a change in rate has occurred with increasing statistical strength of evidence through failure 13. At failure 13, p-v4 (the probability of 4 or more events occurring within an interval equal to the sum of the last 4 TBF values – SUM(C13:C16) - when the mean is cell D16) is 0.0002. If recognized, action could have avoided the last 3 failures. This upper section calculation is made as though MTBF is a parameter. The middle section determines $MTBF_{RV}$ of equation 4. The lower section determines a probability distribution for each p-value, from which p-value confidence intervals are determined.

number of events; however, implementation to realize the full value of the methodology presents a challenge and opportunity. Avoidance of the last three failures in figure 4 required recognition and action at the time of failure 10. The Concorde data was needed by British Airways on the date of crash. The value of the analysis was great on the day of the crash but was of no value at the end of 21 days when risk was at zero. Analysis must be contemporaneous with the failure. Perhaps it is possible by policy and procedure to trigger ad hoc application when the events are so few and so important as commercial fatal accidents; however, the pump is an example where ad hoc analysis cannot be effective. A typical process plant consists of tens of thousands of equipment assets from large turbines to small instruments with several thousands of events occurring every year. In general, neither the asset nor event(s) that trigger a p-value alarm can be known in advance. This necessitates all assets and all events be analyzed with automation so that only the events that trigger probability alarms draw the attention of reliability engineers, maintenance craftsmen, or others depending upon significance. Automation will likewise be appropriate for applying event interval probabilistic analysis to aircraft precursor events to avoid even the first major event.

5 NEW SYSTEM UNRELIABILITY

All the aircraft types in table 1 were introduced into service with low reliability as measured against contemporaries. These systems were, and are being, improved in response to in-service failures and accidents. Only the very earliest failures are seen in table 1. The DC 6 and DC 8 events were tracked over a portion of the fleet life to allow initial low reliability with reliability growth in-service to be clearly seen and understood. The DC 8 was never grounded but the first 13 fatal commercial crashes are evaluated. Cumulative fatal crashes are plotted

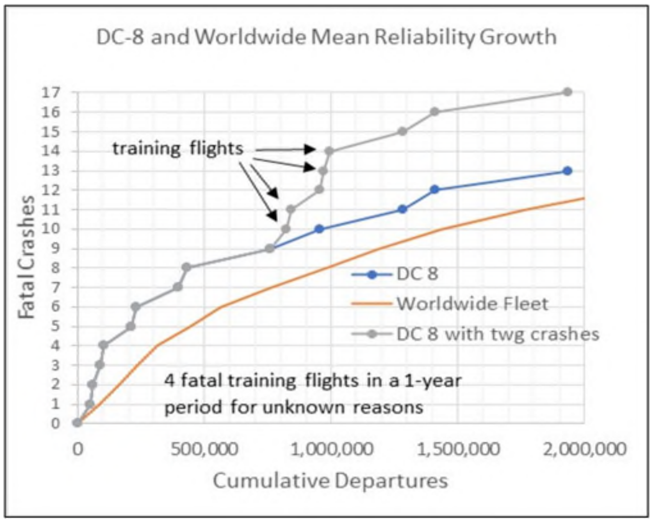

Figure 5

Cumulative crashes versus cumulative departures for the DC 8 and the then current worldwide fleet mean. Both see great reliability growth; however, by the 4th DC 8 crash, the DC 8 had nearly three crashes more that the worldwide fleet mean. This conventional qualitative information is quantified in figure 6.

versus cumulative departures in figure 5. This is a conventional way to see reliability change over time. The worldwide commercial fleet is being fast improved during the time of the 13 crashes. This population mean is also plotted. By the fourth crash, the DC 8 had nearly three crashes more than the population. By the 13th crash, the DC 8 was improved so that it was only about 1.5 events more than the population. The worldwide fleet population MDBE is also improving during this 10-year span, but comparatively, the DC8 initial low reliability is obvious.

Figure 6 is the probability map for the DC 8 with the map changed from that of figure 4 to reflect the worldwide fleet MDBE null hypothesis is a population parameter and that the parameter changes over time. Each p-value is determined with the null that existed at the time of each event.

The DC 6 was grounded after the second event, but the first four events are evaluated. Departure data from 1947 was not found. The domestic air carrier planes in-service and the fatal accidents in 1947[9] provides a null hypothesis of 33,854 mean plane-days between fatal accidents. The dates of the four events[6] and the delivery date for each plane in-service[3] provides the data for figure 7 that show a rate of events that is higher than the domestic fleet. The four events in figure 7 occurred within one year, so the domestic fleet mean does not see reliability growth. It is a single number for that year.

To evaluate figure 7 qualitative data with event interval analysis, the probability map of figure 8 is used. The fleet was appropriately grounded upon the second event. The third and fourth events were failures of the fix installed during grounding. The third event p-values show the plane is still highly unreliable. A p-value of 0.01065 suggest the plane should have been grounded and the fix for the first grounding seriously reevaluated. This should have prevented the fourth event.

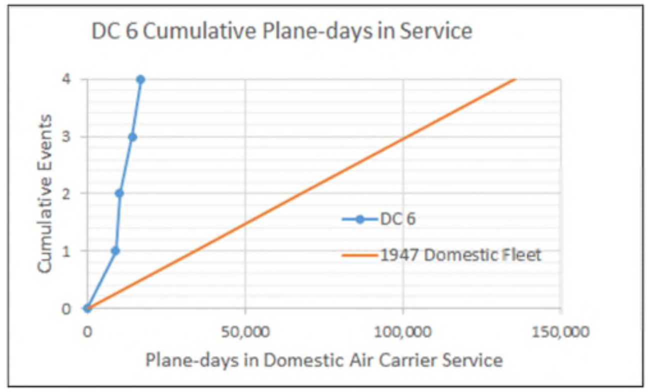

Figure 7

Cumulative events versus cumulative domestic plane-days for the DC 6. The 1st event was not statistically different than the domestic mean, but the next 3 events signaled low reliability that is quantified in figure 8.

6 DISCUSSION

The FAA and aircraft companies are not to be blamed for absence of event interval probability analysis in prior grounding decisions as the null hypothesis to evaluate rate step change is unconventional and only recently published. However, future grounding decisions should include event interval p-values and p-values and their basis should be made public. Phenomenal achievement in aircraft safety has been accomplished, nonetheless, this sampling of aircraft types shows the planes are initially unreliable relative to contemporaries and grounding

=AVERAGE($C$4:C7)

=C9+E8

=1-NORM.S.DIST(((SUM($E$4:E12)-E12)/(B12-1)-E12/2)/(E12*POWER(1/(12*(B12-1)),0.5)),TRUE)

=1-POISSON.DIST(0,SUM(C5:C5)/AVERAGE(D5:D5),TRUE)

=1-POISSON.DIST(3,SUM(C6:C9)/AVERAGE(D6:D9),TRUE)

Column/ Row	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
3	Failure	TBF	MTBF	Cum TBF	LaPlace p-value	p-v1	p-v2	p-v3	p-v4	p-v5	p-v6	p-v7	p-v8	p-v9	p-v10	p-v11	p-v12	p-v13	p-v14
4	1	49420	83333	49420		0.4474													
5	2	7674	76923	57094		0.0949	0.1602												
6	3	32432	76923	89526	0.3210	0.3440	0.0968	0.1062											
7	4	11928	76923	101454	0.1936	0.1436	0.1143	0.0313	0.0423										
8	5	112146	76923	213600	0.8535	0.7673	0.4792	0.3327	0.1679	0.1418									
9	6	14896	125000	228496	0.6583	0.1123	0.3583	0.1899	0.1300	0.0591	0.0533								
10	7	168274	200000	396770	0.9459	0.5689	0.3109	0.3784	0.2567	0.1942	0.1164	0.0984							
11	8	34488	200000	431258	0.8713	0.1584	0.2693	0.1301	0.1790	0.1111	0.0810	0.0452	0.0384						
12	9	326852	222222	758110	0.9910	0.7703	0.5104	0.4698	0.3337	0.3680	0.2889	0.2404	0.1753	0.1522					
13	10	197000	222222	955110	0.9914	0.5879	0.6822	0.4814	0.4508	0.3369	0.3656	0.2974	0.2541	0.1946	0.1716				

Figure 6

DC 8 probability map. Column D is the reciprocal of the worldwide commercial jet fleet fatal event rate at the time of each event. LaPlace p-values use the contemporaneous dataset existing at the time of each event and show improvement over time. Poisson p-values, particularly p-v3 and p-v4 at event 4, give early indication of subpar reliability.

Failure	TBF	MTBF	Cum TBF	LaPlace p-value	p-v1	p-v2	p-v3	p-v4
1	9150	33854	9150		0.23683			
2	1264	33854	10414		0.03665	0.03864		
3	3935	33854	14349	0.18667	0.10973	0.01065	0.00927	
4	2740	33854	17089	0.16628	0.07775	0.01706	0.00180	0.00181

Figure 8
DC 6 probability map. The plane was appropriately grounded upon the 2nd event. Between the 3rd and 4th events, the issue that caused those two events occurred on another aircraft type, but in the absence of alerting p-values, this warning was ignored.[6]

decisions are much worse that those made with a coin flip. Because the analysis spans 75 years, these wrong decisions and initial low reliability are not a function of any one company or organization.

Reliability is improved after introduction into commercial service and after major events have occurred, often with avoidable loss of life. Certification of new aircraft should include the p-value at which a serious event will, in the absence of immediately available cause and effect data showing to the contrary, lead to automatic grounding of the fleet. This performance-based "contract" requirement between the FAA and aircraft companies will provide a driving force for the companies to seek reliability in the design stage, including training, documentation, e.g., that is critical to system performance. Also, aircraft companies as well as carriers will feel pressure to monitor the much less serious precursor events that typically precede a crash or emergency landing. After the DC 10 crash near Paris, it was reported[6] that the fleet had 1,000 incidents with the cargo door in only the prior six months. This was from deterioration in-service of the door locking mechanism. It was a wear out failure mode. The 1,000 incidents would have increased exponentially over time and departures. These precursor events would most likely have triggered probability alarms.

While all six fleets reviewed were placed in-service while unreliable relative to contemporaries, the six fleets are not a random sample. It is biased toward unreliable planes because it is biased toward those that have been grounded. Nonetheless, the results are noteworthy. The FAA and aircraft manufactures can fix the bias issue with event interval probability analysis over a broader range of systems. Furthermore, the FAA, aircraft companies or any researcher are encouraged to analyze fleets in this paper with more precise data. The author's intent is to present data and sources so results can be replicated and improved. The input data for this paper was calculated and estimated from various sources. The details of how the data were obtained and estimated can be found at www.pmfseries.com.

REFERENCES

1. J. Smith, O. Sac, K. Bordelon, "Contemporaneous Failure Time Analysis Using Poisson Probability", *Proc. Ann. Reliability & Maintainability Symp.*, (Jan.) 2018.
2. J. B. Smith, "Failure Time Analysis Applied to Boeing 737 MAX", *Proc. Ann. Reliability & Maintainability Symp.*, (Jan.) 2020.
3. S. R. Lynn, Douglas Production List of DC8- DC9-DC10, Middlesex, U.K., Airlines Publications, 1977.
4. Boeing, "Statistical Summary of Commercial Jet Airplane Accidents – Worldwide Operations, 1959 - 2017", 2017.
5. NTSB Accident Report, American Airlines, Inc. DC-10-10, N110AA, Chicago-O'Hare International Airport, Illinois, May 25, 1979.
6. P. Eddy, E Potter, B. Page, *Destination Disaster*, New York, Times Newspapers, Ltd., 1976.
7. NTSB Accident Report, American Airlines, DC-10-10, N103AA, Near Windsor, Ontario, Canada, June 12, 1972.
8. Bureau Enquetes-Accidents, Accident on 25 July 2000 1at La Patte d'Oie in Gonesse (95) to the Concorde registered F-BTSC operated by Air France, 17 July 1978.
9. U. S. Department of Commerce, *Statistical Handbook of Civil Aviation*, 1948.
10. Japan Transport Safety Board, Emergency Evacuation Using Slides All Nippon Airways Co., LTD. Boeing 787-8, JA804A Takamatsu Airport at 08:49 JST, January 16, 2013.
11. A. Tangel, A. Pasztor, "Regulators Found High Risk of Emergency After First Boeing MAX Crash", Wall Street Journal, Dec. 11, 2019 (link to FAA draft document, since removed).
12. J. B. Smith, "Boeing 737MAX Thru DC6 Fleet Grounding Decisions Revisited with Event Interval Probability Analysis", *Proc. Ann. Reliability & Maintainability Symp.*, (Jan.) 2021 (To be published).

ACKNOWLEDGEMENT

The event interval probabilistic analysis methods and data analysis tools and processes were developed by Zachry Industrial's Maintenance and Reliability Services Group to support process industry reliability. The methods and tools would not exist but for their progressive thinking and support.

BIOGRAPHIES

Jan B. Smith, PE
Reliability Engineering Consultant
e-mail: janb.smith@cebridge.net

Jan B. Smith, PE, is a Reliability Engineering Consultant with a 55-year career in reliability. He has established and managed reliability departments within major corporations and independent professional engineering firms. Now a private consultant, he has significant experience in root cause analysis; finite element analysis; statistical and probabilistic analysis; capacity and availability forecasting using empirical probabilistic methods, holding three patents on this method. He has chaired conferences on reliability, authored technical papers on the subject, and has developed and taught in house and public seminars on root cause analysis. He is a registered Professional Engineer.