Boeing 737MAX Thru DC6 Fleet Grounding Decisions Revisited with Event Interval Probability Analysis

Jan B. Smith, P.E., Reliability Engineering Consultant

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

SUMMARY & CONCLUSIONS

The application of statistics and probability to event timing data is an unrecognized powerful decision-making aid. 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. Event interval probability analysis is independent from cause and effect and organizational and other interfaces, e.g., human factors design and pilot error. It considers bottom-line total system performance only.

Statistics and probability analysis demand much data; however, for serious critical events failure 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 perfect 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, and for this paper, the number of departures between accidents are used, except where noted.

This paper reviews six fleet groundings on five aircraft types with 13 different grounding decisions. Data for the analysis and all analysis results are presented; however, the first opportunity to ground decision is the most important. The first opportunity to ground decision is retrospectively judged to be wrong if future events unfolded that demonstrate earlier grounding would have been appropriate.

Decisions for five of the six fleet groundings revisited involved US designed and manufactured aircraft with grounding decisions made by the FAA or its predecessor organizations. On these five critical decisions, the grounding of the DC 10 due to a crash from engine pylon cracking on 5/25/79 was correct. The other four decisions were proven to be incorrect by future events. The FAA made only one correct grounding decision out of five.

Event interval probability analysis failed to ground on first opportunity upon a DC 6 crash on 10/24/47, using data that existed contemporaneously with the crash. The other five first opportunity decisions (including the Concorde) had p-values less than 0.025. The method would reject the null hypothesis and ground on these five occasions. So, the application of event interval probability analysis would lead to five out of six correct decisions, in the total absence of cause and effect data.

There is always a possibility of a false positive with these statistical methods. The total false positive probability for the five correct decisions is 0.0467 for an average false positive per grounding decision of 0.0093. An average one percent chance of unnecessarily grounding a fleet is small relative to the risk of not grounding timely, as the current Boeing 737 MAX situation demonstrates. The first crash of the 737 MAX has a p-value of 0.022. Upon the second crash, the p-value is 0.00099. Grounding by p-values would have led to immediate grounding upon the first crash, and certainly after the second.

When the null hypothesis is rejected, the alternative hypothesis that the fleet accident rate is above expectation by a statistically significant amount is accepted. Upon acceptance of the alternative hypothesis, unreliability probability distributions are obtained via computer simulation for the demonstrated low reliability aircraft fleet. From these probability distributions, the risk of continuing to fly the low reliability aircraft is found. For example, the risk of flying three days following the second 737 MAX crash, assuming all the fleet was flying, was 4.75% chance of a third event. After the Concord crash, British Airways flew an additional 21 days following the crash with a significantly low reliability aircraft and incurred an unrecognized 1.12% chance of a second event.

The current FAA and predecessor decision record of 20% correct versus 83% correct using event interval probability analysis indicates that the FAA and aircraft manufactures can improve their decision-making by incorporating the method. Note this poor decision record covers 75 years; therefore, cannot be attributed to current organizations and individuals. The following specific steps are recommended:

1- The FAA should conduct an event interval probability analysis of the aircraft events in this paper using best available data and publish results.

2- The method should be conducted immediately upon future major events, such as crashes, by manufacturers or the FAA. Departure intervals and p-values should be made public.

3- Air worthiness certification should specify the p-values at which a serious event will lead to automatic grounding of the fleet, in the absence of immediately available and strong cause and effect evidence indicating to the contrary.

4- To help avoid even the first accident, the FAA and/or manufacturers should continuously monitor accident precursor event p-values. These precursor events typically precede a more serious event, e.g. maintenance and operational issues. This will require automating the analysis due to volume and complexity.

1 INTRODUCTION

The analysis methodology, here applied to aircraft fleet grounding decisions, was first published in 2018[1] with application principally in process industries. It was applied to the Boeing 737 MAX in 2020[2]; therefore, it is unrealistic to expect the FAA and Boeing to have applied the method to the 737 MAX decisions or any earlier decisions. The purpose of this paper is to assure the method is used by the FAA and aircraft manufacturers in the future.

Traditionally, event interval analysis is applied to datasets to determine if there is a trend in the dataset and to forecast the number of future events. Here analysis is applied to individual events and contiguous groups of events for identification of a step change or shift from expected rate. The objective is immediate recognition of the step change so intervention can correct the issue and thereby avoid future events. The entire dataset is not used except perhaps to establish the expectation by way of the mean for the system. The emphasis is on preventing future failures; therefore, analysis must be contemporaneous with the event. The value in the analysis is not historical after the fact information. The value is in being alerted that an event has triggered a p-value denoting need for intervention when otherwise the significance of the event interval would be ignored or underappreciated.

Data for this analysis was pulled from many sources and estimates were required. Raw data and results of the analysis are found in table 1. Table notes reference data sources and basis for estimates.

The FAA and/or aircraft manufacturers should conduct their own event interval analysis with the more precise data that assuredly are obtainable. Furthermore, while no aircraft was excluded because it did not fit a narrative, there was no attempt to obtain a statistical sample of crash events or aircraft types. It may be appropriate for the FAA to expand the analysis to other systems and report results and show method.

2 ANALYSIS

The results of all events and aircraft types are presented, but Boeing 737 MAX data are selected to present the methodology. Data are analyzed with three concepts:

- Poisson Probability
- Computer Simulation
- Statistical Process Control (SPC)

These three concepts are completely different with apparently nothing in common. The use of three concepts are to promote understanding and acceptance. Analysis results are harmonious and nearly identical. This is because the underlying mathematics are the same. Poisson probability and simulation are both required to fully implement the methodology in practice.

Throughout this paper, decimal places beyond those that are truly significant are used. For example, probability values may be shown to 5 decimal places for the purpose of allowing the calculations to be duplicated by others for understanding, applying better data, and for use in other applications.

2.1 *Poisson Probability and Null Hypothesis*

Poisson probability is used to determine the probability of specific numbers of events occurring within a specified time interval, when the events are generated by a homogeneous Poisson process (HPP). Failures times are independent and identically exponential distributed random variables. The expected number of events are constant for any time interval of equal length. Repairable system failures are, in general, such an HPP. But new failure modes, improper repair, and any other special cause produces failure times that do not fit the HPP conditions for Poisson. Moreover, it is these nonconforming special cause failures that are of most interest. Therefore, on the surface, using Poisson to find special cause failures that do not conform to the requirements of Poisson use may appear to be inconsistent. But here the Poisson is used in reverse to identify time between failure (TBF) data that appear not to conform to Poisson distribution requirements.

The Poisson probability of events, with time being the Poisson interval, is:

$$P(x) = (e^{-\mu})(\mu^{x})/x! \text{ and} \tag{1}$$

$$\mu = \lambda t = t/MTBE \tag{2}$$

Where:

P(x): The Poisson probability that exactly x events are experienced, given the mean is μ.

x: Specific number of events in a specified time interval, x = 0, 1, 2, 3 ….

μ: The mean number of events expected in a specified time interval.

e: An approximately 2.71828 constant, the base for natural logarithms.

λ: Event rate, number of events over time.

t: time interval

MTBE: Mean time between events.

For aircraft accidents, the appropriate Poisson interval is not time, but number of departures (flights or cycles) for the aircraft type. The common time between failure (or event) is replaced with departures between events (DBE) with mean departures between events (MDBE) determined by the worldwide commercial jet fleet MDBE existing at the time the event being analyzed was experienced.

The first Boeing 737 MAX crash occurred upon 135,980 departures[11]. Departures between the first and second crash are unknown to the author, but in an earlier paper[2], those departures were conservatively estimated to be 139,313 and this estimate is used. Also, the mean (MDBE) for the worldwide fleet was determined to be 6,105,714 departures between fatal accidents (35 fatal accidents in 213.7 million departures)[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 while p-v2 is the probability of two or more events occurring in the sum of the last two DBEs, as described in detail in earlier papers [1, 2, 12].

1st crash: p-v1 = $P(x \geq 1) = 1 - P(x=0) = 0.02202$

2nd crash: p-v1 = $P(x \geq 1) = 1 - P(x=0) = 0.02256$
p-v2 = $P(x \geq 2) = 1 - [P(x=0) + P(x=1)] = 0.000986$

The null hypothesis that the 737 MAX is as reliable as the worldwide fleet mean is rejected upon the first crash and is overwhelming rejected at the time of the second crash. We accept the alternative hypothesis that the system is less reliable than the worldwide mean and the difference is statistically significant. The quick events are not simply random variation events of an otherwise reliable system. P-values for both Poisson and computer simulation are reported in table 1 for all accidents in the analysis.

2.2 *Simulation Probability – Null Hypothesis Evaluation*

Probability values for testing the HPP null hypothesis can also be obtained by computer generated DBE samples. This is useful in promoting understanding and acceptance of the event interval methodology. Furthermore, computer simulation is required to assess risk of operating after the null should have been rejected but was not. Equation 3 is the cumulative probability of failure for the independent and identically distributed exponential failure time of a repaired like new and steady state repairable system – our null hypothesis. Each random number generates a sample failure time. Equation 4 changes the Poisson interval of time to number of departures, as this measurement is more appropriate for aircraft accidents.

$$F(t) = 1 - e^{-t/MTBF} = RN \qquad (3)$$
$$RN = \text{Random Number}$$

With departures replacing time, equation 4 provides a DBE sample (DBE_s) for each random number.

$$DBE_s = -MDBE*(\ln(RN)) \qquad (4)$$

Our null hypothesis and the conditions for which equations 3 and 4 are valid are the same. Equation 4 produces a large number (one for each random number) of DBE samples from a population with a known mean. The parameter mean can be a goal or target. For the airplane systems analyzed here, the reciprocal of the worldwide commercial jet fleet fatal accident rate at the time of the accident is used. For the DC 6, it is in-service plane-days divided by fatal domestic accidents.

Applying equation 4 to the 737 MAX, numerous DBE samples from the worldwide fleet with MDBE 6,105,714 forms the probability distribution in figure 1. The probability of a crash occurring by 135,980 departures is the proportion of area under the curve to the left of that value. This probability is 0.022 and is consistent with the Poisson p-v1. This low p-value should reject the null hypothesis.

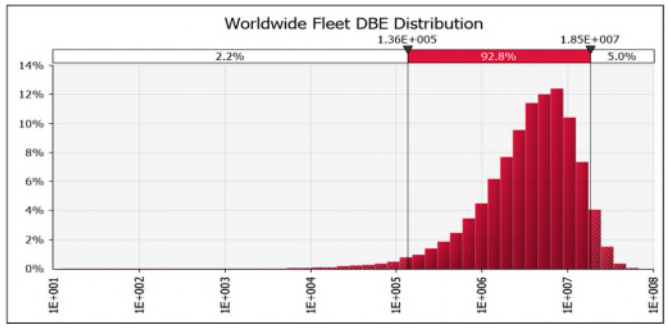

Figure 1
Probability vs DBE distribution of the DBE samples from the worldwide fleet using equation 4. The probability of experiencing an event within the number of departures of the first 737 MAX crash by random chance is 0.022.

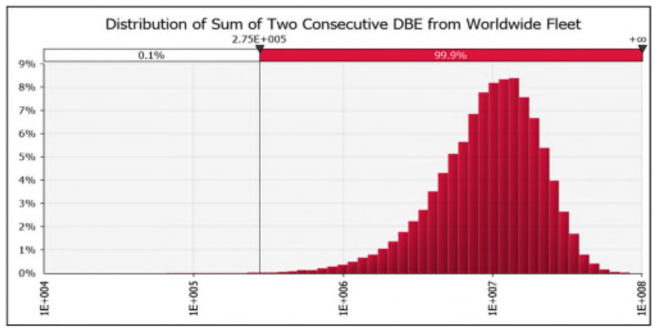

Figure 2
Probability vs DBE. Two consecutive events within 275,293 departures will occur about one time in a thousand within the worldwide fleet. Therefore, the null hypothesis is rejected. The 737 MAX is less reliable than contemporaries with statistical significance.

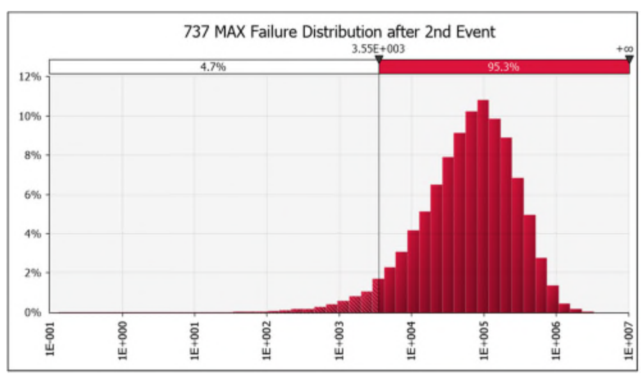

Figure 3
Probability vs DBE distribution upon the 2nd 737 MAX crash, or unreliability distribution. This shows the probability of a crash within any certain number of departures. The probability of a crash within 3,948 departures[2], the estimate for all planes flying for 3 days, is 0.0475.

Now Equation 4 is used to obtain the sum of two consecutive random draws from the worldwide fleet population. The distribution of this sum is seen in figure 2. In the worldwide fleet population, the probability of two consecutive events occurring within a total of 275,293 departures is 0.00099. In other words, the worldwide fleet may see two events or more within this interval as often as one in a thousand by random chance. Using the logic of classical statistics, we should reject the null and accept the alternative hypothesis.

The aircraft is proven via statistics and probability to be unreliable relative to the goal, target or standard. In this case the standard is the worldwide fleet fatal accident rate. The risk of a false positive is the p-value. The probability of an inappropriate grounding, a false positive, upon the second crash is 0.00099 – determined with data available at the time.

With the null hypothesis rejected, it is of interest to measure the risk of continued operation of the unreliable aircraft. The 737 MAX MDBE is unknown. We have two DBE data samples from a population with an unknown true mean. This dataset mean, 137,647, is used with equation 4 to generate numerous samples of the mean of two consecutive DBE values. MDBE in equation 4 is now a random variable. Equation 4 is used again, but within the same calculation, with a different random number draw to generate DBE samples. Numerous samples form the probability distribution of figure 3. This is the departures to fatal accident distribution, or unreliability distribution. The area under the curve to the left of the departures acquired in 3 days is the risk of a third event. The process of using equation 4 twice to obtain MDBE as a random variable, then again using the random variable mean sample to obtain DBE samples to failure is described in more detail in reference 12.

The plane was grounded by Boeing and the FAA three days after the second crash. The probability of a third event within three days, if all planes were flying, is 0.0475. The risk exposure in the three days is 8.22 virtual fatalities (Risk=Probability * Consequence or 0.0475 * (189 + 157)/2). The risk of not grounding timely can be under appreciated because of the long left-tail of the unreliability distribution of figure 3. Note the log scale of the distributions that partially obscures the length of the tail.

2.3 *Statistical Process Control*

Many engineers and others are familiar with statistical process control and quality control charts. Seeing the data within that concept may aid in appreciating the null hypothesis and its rejection. Figure 4 shows 350 simulated DBEs from the worldwide fleet population placed on a statistical control chart. The DBE distribution for the worldwide fleet is seen on the left of figure 4. This distribution is identical to figure 1 rotated 90 degrees. The 350 control chart points can be thought of as having been drawn from the distribution. The two 737 MAX events fall into the out of process control region by control chart rules. The p-v2 value is not defined in the control chart concept due to granularity; however, its Poisson and simulation value is appropriately placed and is just outside the -3sigma control line. If granularity could be improved, the SPC concept would converge with the Poisson and simulation concepts as granularity became finer. The basic mathematics

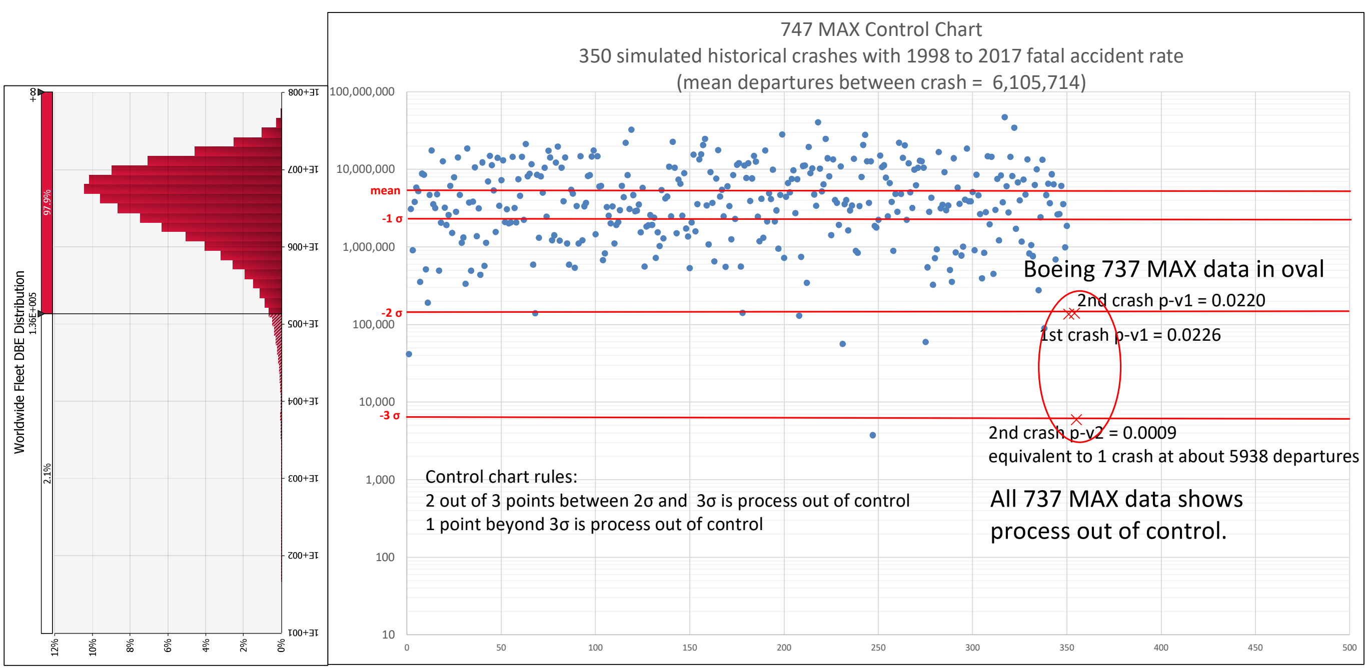

Figure 4

Statistical process control chart for 350 simulated DBE from the worldwide fleet. The distribution for this fleet is on the left. All the simulated events are random variation. By chance some are outliers. The 737 MAX two event are just outside the -2σ line. By control chart rules the process is out of control. The p-v2 probability is identical to one event at 5938 departures as measured by p-value. This point placement is just outside the -3σ line. This point is also in the rejection region.

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. Correct grounding decisions are green and incorrect are red. Yellow means correct decision but late with incurred risk of additional fatalities. Virtual fatalities are the expected value from risk = probability * consequence.*

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		
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		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. 7, new procedure introduced on accident aircraft 55 days earlier with 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 7 page 146.
18- Reference 6.
19- details on estimates can be found at www.pmfseries.com

behind the three concepts are the same.

2.4 *Individual Aircraft Analysis*

Analysis by aircraft type has been discussed so far. Two occasions were identified in which individual aircraft data analysis would have been helpful. On 6/12/72 a DC 10 required an emergency landing due to a door locking mechanism failure that allowed a cargo door to blow open, leading to a crash landing. From the mechanism, prior events, and human interfaces description(6), it appears to be wear-out failure mode at the subsystem level. This is wear-out in the sense of a subsystem deteriorating with time in service with a Weibull failure distribution shape parameter greater than one. The aircraft was only 46 days since delivery, but still one of two of the oldest aircraft in service. The shape parameter is unknown, however, if a value of one is used as a conservative estimate, the Weibull failure distribution reduces to equations 3 and 4. The resulting p-value of 0.00128 is likely larger than actual. The event was treated as though it were a random event on a reliable system, not as proof that the system was highly unreliable. The fleet p-value seen in Table 1 is 0.01686. The fleet could have been grounded whether by fleet or individual aircraft p-value. It was not grounded and not reliably fixed and twenty months later the same door locking system caused 346 fatalities.

The DC 10 crashed on 5/25/79 due to cracking of a pylon attaching an engine to the wing. Cracking was initiated due to a maintenance procedure change on a few aircraft a short time earlier(5). In this case DBE are measured from the procedural change and only for those aircraft exposed to the new procedure. The p-value is 0.00016. The fleet was grounded, obviously without event interval probability analysis, and is the only correct FAA first opportunity to ground decision of those reviewed.

3 SUMMARY RESULTS & RECOMMENDATIONS

Table 1 contains the raw data and analysis results for the various airplane types and events analyzed. All p-values that are in red font indicate values below 0.025 that should reject the null hypothesis. All the aircraft types are shown to have been unreliable relative to the then worldwide fatal accident rate, using data for the analysis that existed at the instant of the event. Where actual data were unknown to the author, estimates were derived from other data in various ways. See reference 2 as an example. The FAA and aircraft companies have access to more accurate data. They should conduct this analysis with data as exact as possible.

The FAA and aircraft companies are not to be blamed for the absence of event interval probability analysis in grounding decisions as the null hypothesis to evaluate rate step change is unconventional and only recently published. However, future grounding decisions should consider event interval p-values and p-values 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. Reliability is improved after introduction into commercial service and after major events have occurred, often with avoidable loss of life. (Reference 12 expands the aircraft type sample and underscores this point). 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 safety. Also, aircraft companies as well as carriers will be incentivized 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 that the fleet had 1,000 incidents with the cargo door in only the prior six months(6). The precursor events most likely would have triggered probability alarms.

The use of p-values in grounding decisions means there is a chance of a false positive that will unnecessarily ground a fleet. The average false positive seen in Table 1 is 0.00934. This is an extremely small risk relative to, for example, the impact of the 737 MAX not being grounded upon the first crash. Also, the unrecognized high risk of a third event exposed Boeing and the FAA to exponentially more severe impacts. The use of event interval p-values will benefit the aircraft companies, the FAA, and the flying public.

REFERENCES

1. J. Smith, O. Sac, K. Bordelon, "Contemporaneous Failure Time Analysis Using Poisson Probability", *Proc. Ann. Reliability & Maintain 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 at 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. Smith, J. B., "Failure Interval Probabilistic Analysis for Risk-based Decisions – Concorde Crash Example", 38th International System Safety Conference, Nov 2020.

ACKNOWLEDGEMENT

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
3342 Courtland Manor Ln
Kingwood, TX 77339
e-mail: janb.smith@cebridge.net

Jan is a Reliability Engineering Consultant with a 55-year career in reliability. He 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, and statistical and probability analysis. He developed system capacity and availability empirical probabilistic methods leading to three business method patents. Jan has chaired conferences on reliability, authored technical papers on the subject, and has developed and taught client specific and public seminars on root cause analysis. He is a registered Professional Engineer. website: www.pmfseries.com.