Avoiding System Failures with Event Interval Probability – 737 MAX Case Study

Jan B. Smith, PE

Key Words: Probability values, p-values, Poisson, null hypothesis, prior probability, Monte Carlo, risk assessment, reliability distributions, event interval probability, EIP

SUMMARY & CONCLUSIONS

A newly developed data analysis methodology drawing upon statistics, probability and reliability theory and Monte Carlo simulation uses event interval probabilities to recognize real reliability degradation in a system with random variation in the data. The analysis method uniquely allows a complete probabilistic analysis with a dataset as small as one; therefore, it is useful for systems requiring high reliability where system failures must be few. It has been shown capable of avoiding failures such as the Boeing 737 MAX crashes with the 346 fatalities and tens of billions of dollars in economic loss [1,2,3,8]. The purpose of this paper is to encourage use of the analysis method in the future – not only in commercial aviation but other applications. The 737 MAX datasets provide an opportunity to illustrate both the method and value of the data analysis.

Updated datasets for the Boeing 737 MAX crash events and precursor events are analyzed and results presented. Aspects of the analysis methodology applied to the 737 MAX have been presented in various papers since 2020 [1, 2, 3], using the best data and estimates that were assembled at that time. Here we consolidate the applicable elements of the analysis method with results in one paper, using current best publicly available raw data for the analysis. The event interval probability (EIP) analysis input data used is that which was available at the time of the crash or precursor. A precursor event is a relatively minor event that can be an early warning of an impending major event such as a crash. The 737 MAX precursor is "sustained stick-shaker activation".

The unreliability of the 737 MAX is identified upon the first crash with a probability of 0.02202 that the event could be random variation of a fleet that is as reliable as the worldwide fleet. This p-value rejects a null hypothesis and unreliability distributions – departures to crash distributions – allows risk to be quantified. The distributions are formed with Monte Carlo simulations to solve a math problem – not computer simulation in the normal sense of evaluating a model. The probability of a crash within the nine days to an air worthiness directive incurred a risk with expected value of approximately 23.0 fatalities.

Upon the second crash, the probability of the events occurring in such a short number of departures by random chance is 0.00084. With the null hypothesis rejected, the risk associated with continued operation of the unreliable fleet for three days before grounding, considering all the fleet flying, is about 7.9 fatalities. This high risk is due to a currently unrecognized long left tail of the unreliability distribution. Because of high fatality risk within even short intervals, the risk of uncorrected continuation in-service should be quantified immediately upon any serious event with low p-values.

The first sustained stick-shaker activation precursor occurred on October 28, 2018. Input data for this precursor is calculated from publicly available sources. EIP analysis could recognize the event as a statistical outlier and provide an opportunity to avoid even the first crash. The probability of the precursor is 0.0235 before a Bayesian-like prior probability is applied. The final probability adjusted for the prior information is 0.0136. This low p-value would alert to the event interval being a statistical outlier signaling need for investigation. This would provide another opportunity to recognize a possible reliability issue, ground the problem plane, and investigate. If action on this precursor had been taken, it is possible to have avoided the first crash. An appreciation for the importance of precursors is a prerequisite to their analysis.

1 INTRODUCTION

Significant events on important systems often are very few because of high reliability and safety demands; however, it is typical that statistical and probabilistic analysis need a quantity of data. The event interval probability (EIP) methodology is as applicable to a dataset of one as to a dataset of one hundred. Also, the method is unchanged with any dataset size. This makes EIP especially valuable in situations where high reliability and high safety are required – where even one event is one too many. The data analysis and risk assessment method comprising EIP is demonstrated with the Boeing 737 MAX crash and precursor events. All analysis input data come from or are derived from public sources.

A null hypothesis is taken that the dataset is generated by a homogeneous Poisson process. This allows the Poisson distribution to be used to calculate Poisson probability values for all the dataset event counts within the intervals associated with those counts. As the Poisson probability approaches either zero or one, it is statistical strength of evidence of reliability degradation or improvement, respectively. Sufficient strength of evidence rejects the null. After the null is rejected, risk of the uncorrected system remaining in-service is determined with Monte Carlo generated probability distributions for system unreliability.

1.1 *Null Hypothesis*

A null hypothesis is taken that the system producing the event data is a homogeneous Poisson process (HPP) with a mean interval equal to that of a comparative population mean. For the 737 MAX, this population is the worldwide jet commercial scheduled carrier fleet. Poisson probability is used, but not directly. Event data of interest will not meet the conditions for using Poisson. However, Poisson can be used in reverse by applying event data to the null hypothesis. Poisson p-values tell us how well our event data matches the comparative population. These p-values provide strength of evidence to reject or not reject the null hypothesis. P-values are calculated for individual events and contiguous groups of events. This is to identify an event interval or group of event intervals that signal a statistically significant deviation from random variation within the comparative population.

1.2 *Poisson Probability*

The Poisson probability of any number of events occurring within any Poisson interval is given by equation 1. The compliment of equation 1 will calculate the probability that our dataset fits the Poisson distribution. If our dataset fits poorly, indicative of a low probability, we will reject the null. The calculated Poisson probability will be used like a common p-value that evaluates a test statistic.

$$P(x;\mu) = (e^{-\mu})(\mu^{x})/x! \qquad (1)$$

Where:

$P(x; \mu)$: Probability that exactly x events occur within a specified interval when the expected number in the interval is μ

x: Specific number of events in an interval, x = 0, 1, 2 ….

μ: Expected number of events within an interval

e: Euler's number.

For the null hypothesis that is an HPP, the expected number of events over an interval, μ, is developed below and with a Poisson interval of time. Time is a common Poisson interval and is more likely to be familiar to the reader.

$$\mu = \lambda t \qquad (2)$$

$$\lambda = 1/MTBE$$

$$t = \sum TBE$$

$$\text{therefore:} \quad \mu = \sum TBE/MTBE \qquad (3)$$

Where:

λ: Expected event rate, event counts per time

t: Time between event (TBE) or $\sum$TBE

MTBE: Expected (Mean) time between events

$\sum$TBE: Dataset sum of time between contiguous events (further explained by examples)

2 *BOEING 737 MAX CRASH INTERVAL*

For evaluation of aircraft crash events, departures are the better Poisson interval with departures between event (DBE) and its mean (MDBE) used for TBE and MTBE, respectively. For the null hypothesis to which we apply the above equations, there is no distinction between the first event interval and subsequent intervals. In some data analysis contexts, departures to first event would be appropriate, but here we can use the term "departures between event" for even the first event.

2.1 *First crash*

The first crash on 10/29/2018 occurred with a 737 MAX fleet DBE of 135,980 departures [4] with 6,105,714 worldwide fleet MDBE calculated from published data [5] - number of fatal crashes over a recent 10-year period and the number of departures over the same period. With Poisson probability for the first event being defined as event 1 p-v1, we get the probability of one or more events within the interval of the one event using the complement of equation 1.

$$\text{p-v1} = 1 - (e^{-\mu})(\mu^{x})/x! = \underline{0.02202}$$

where x=0

DBE = 135,980

MDBE = 6,105,714

μ = DBE/MDBE = 135,980/6,105,714 = 0.02271

Event 1 p-v1 answers the question, what is the probability of one or more events occurring within the interval of the one event? This Poisson probability is used like a common p-value as strength of evidence against the null. This low p-value indicates how poorly our data fits the null hypothesis. The data is not likely to be random variation in a fleet that is as reliable as its peer group, the worldwide fleet. This p-value is sufficiently low to reject the null and accept the alternative hypothesis that the 737 MAX fleet is unreliable compared to the peer group.

2.2 *Second crash*

The number of departures to the second crash in not believed to be publicly available and is estimated from data that are available. Plane delivery dates are available from the Boeing website. The delivery date for individual planes allows plane-days in service to be aggregated for the fleet. The first crash plane-days in service is 47,711 and to the next is 41,352. From the departures to the first crash and the plane-days to the first crash, the average number of departures per day is 2.85. Using this first interval average as an estimate for the second interval gives 2.85 * 41,352 = 117,856 departures.

Using equation 1 and 3, the second event p-v1 (the probability of the second event occurring in an interval of the second event) is:

$$\text{p-v1} = 1 - (e^{-\mu})(\mu^{x})/x!) = \underline{0.01912}$$

where:
$x=0$
DBE = 117,856
MDBE = 6,105,714
μ = 117,856/6,105,714 = 0.01930

Using equation 1 and 3 with departures as the Poisson interval, the second event p-v2, the probability of two or more events within the interval of the two events, is:

$$p\text{-}v2 = 1 - \{(e^{-\mu})(\mu^{x0})/x0! + (e^{-\mu})(\mu^{x1})/x1!\} = \underline{0.00084}$$

where:
$x0 = 0$
$x1 = 1$
$\mu = \sum$DBE/MDBE = (135,980 + 117856)/ 6,105,714 = 253,836/6,105,714 = 0.04280
MDBE = 6,105,714

The probability of two or more crashes within the interval of the two crashes, if the fleet is as reliable as the peer group, is 0.00084. The approximately eight chances in 10,000 that the 737 MAX is as reliable as the worldwide fleet is overwhelming evidence for rejecting the null. The alternative hypothesis that the fleet is less reliable than its peers is accepted.

But before the null is rejected and we analyze the risk of not correcting the problem, we consider more deeply what p-value is and how it can be changed to even lower values.

3 POISSON VERSUS COMMON P-VALUES

In statistics a p-value is commonly associated with a test statistic. The common p-value informs as to the likelihood of the test statistic being as extreme, or more so, than the value produced by the dataset. This p-value is not related to the dataset directly. It does not indicate the probability of the null being true; however, there is a common misconception that it does.

The Poisson p-values are related to the dataset – not to a test statistic. This p-value is the probability that the null is true. It is not the misconception associated with common p-values. Table 1 illustrates the difference between the two p-values – common and Poisson.

Common p-values	Poisson p-values
Shows how rare the test statistic is given the null hypothesis is true	Shows how rare the data are given the null hypothesis is true
Based on a model (like Z or t statistic)	Based on mathematics - no model beyond use of a null hypothesis
Uses assumptions	No assumptions
Uses variation within data (like variance)	Does not use variation within the data
Uses number of samples	Does not use number of samples
P-value is not the probability of the null being true	P-value is the probability of the null being true

Table 1 – Poisson Versus Common P-values

While common p-values are derived from frequentist statistics and are not changed by prior information or belief, mathematically derived Poisson p-values can be modified by a prior probability.

4 PRIOR PROBABILITY

As with Bayesian statistics, the Poisson probability can be changed by a prior probability that reflects prior information or belief. While the term prior or prior probability is associated with Bayesian statistics, Bayesian statistics is not used. We only borrow from Bayesian statistics the concept of using a prior probability and combine that with Poisson probability.

An example of why we sometimes need to use the prior is illustrated by the 737 MAX input data to equation 3, using DBE and MDBE in lieu of TBE and MTBE, respectively. MDBE is the average number of departures between crash events for the worldwide fleet that consists of demonstratively reliable aircraft. The 10-year period from which the mean was calculated preceded the 737 MAX crashes and included no crash of a new aircraft fleet design. DBE in equation 3 is from a new design type that is not mature and does not have proven reliability. If the MAX was a mature fleet, then the p-values could not be improved with this prior, but being a new design provides additional information.

Table 2 showing design flaw history provides the basis for establishing the prior probability that can be multiplied by the 737 MAX p-values. Eleven of 19 designs did not have a design flaw so severe that correction was required. The probability of any new fleet design being acceptably reliable can be taken as 11/19 or 0.579 probability. This prior probability, which is

Plane Type	Significant Flaw Requiring Correction?	Flaw Description	Flaw Caused Fatalities?	Flaw Failure Rate
DC 6	Yes	Inflight fuel spillage & fire	Yes	Constant
DC 6	Yes	CO2 in cockpit	Yes	Constant
DC 7				
DC 8	Yes	Cockpit human factors	Yes	Constant
DC 9				
DC 10	Yes	Cargo door latching system	Yes	Increasing
MD-11				
MD-80				
MD-90				
707				
717				
727				
737	Yes	Rudder control	Yes	Constant
747				
757				
767				
777	Yes	Engine blading fatigue		Increasing
787	Yes	Battery fire		Constant
737 MAX	Yes	Loss of control	Yes	Constant

Table 2 – Flaws in new design types discovered in-service and sufficiently severe that correction was mandatory. Eleven designs out of 19 were reliable when placed into service (DC 8, 737 & 777 design flaws discovered later in life). The Bayesian-like prior probability is 11/19 = 0.579.

independent from the p-values, can be multiplied by the Poisson p-values to get lower adjusted p-values that use all available information.

Interpretation of Table 2 is subjective, and the use of a prior probability is sometimes criticized because of this nonscientific feature of Bayesian statistics. Some intelligent people could see the history of new designs in-service as irrelevant to the reliability of a particular new design that is in-service. For these people, the prior probability is 1.0. The prior will have no impact on the final p-value. Other intelligent people may think that only recent history is important and use the last three designs with zero out of three chance that the new design does not have a flaw that will require correction when found. Their prior would be 0.0 probability that the new fleet is acceptably reliable. This range of results for the prior is the total possible range. It extends from zero to one probability; therefore, the basis for the prior needs to well considered and explained. The prior should be used with caution, as it can lead to bad results, while not using it when needed is also wrong. Choosing to use the entirety of Table 2 for this paper, the prior probability of 0.579 can be multiplied by any 737 MAX Poisson p-values to improve the strength of evidence for null rejection. The 737 MAX event Poisson p-values are so low that they reject the null without the prior being needed, but the prior will be applied to the precursor Poisson p-value to demonstrate method.

5 *PROBABILITY DISTRIBUTIONS / RISK ASSESSMENT*

With the null rejected and the alternative hypothesis that the fleet is not as reliable as the worldwide fleet accepted, the departures to next crash distribution is needed. This distribution informs as to the risk of continued operation without correction. It provides the probability and fatality risk.

Figure 1 describes the basic process for Monte Carlo simulation to obtain departures to crash probability distributions. The cumulative failure distribution, equation 4, is the complement of reliability, or unreliability. This is the complement of equation 1 with x = 0.

$$F(t) = 1-e^{-\lambda t} \qquad (4)$$

Failure rate λ in equations 1 and 4 is constant. Before the null is rejected, the null hypothesis assures the constant failure rate by definition – it is inherent to the null hypothesis. Now that the null is rejected, use of equations 1 and 4 requires an assumption of a constant failure rate. That assumption must be justified.

The null hypothesis is that the 737 MAX fleet is at least as reliable as the worldwide fleet peer group. This null hypothesis 737 MAX fleet is defined to be system A. Once the null is rejected, the 737 MAX fleet is system B. System B is unreliable compared to system A. The difference in the systems is “problem causes”; otherwise, these two systems are identical. These problem causes do not need to be known, and they are usually unknown or at least not fully understood. The failure events that rejected the null can be considered a dataset from system B. Initially, system B has only one crash event, and later has two events. It would require several crash events to confirm with failure data that system B has a constant failure rate, but the very purpose of EIP is to eliminate these future failures by acting as soon as p-values signal unreliability, hence the requirement of an assumption.

The basis for the constant failure rate assumption applied to this system B (the 737 MAX fleet with the alternative hypothesis accepted) is:

- Theoretically, complex repairable systems have a constant failure rate. System B has a higher failure rate, but it is expected to be constant from reliability theory.
- By engineering judgement, the historical design flaws in commercial aircraft types in Table 2 are mostly constant failure rate. Use of a constant failure rate in the two cases of increasing failure rate is a close approximation and provides a lower bound on risk.
- EIP experience with many datasets demonstrates constant failure rate until the system B is corrected by intervening and implementing change. When the causes of reliability degradation are eliminated, system B reverts to system A, as is logical.

Given the constant failure rate assumption, a uniformly distributed zero to one random number is transformed to a time sample, or DBE sample using Equation 4. Equation 4 is set equal to a random number (RN) and the equation solved for t.

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

RN = uniformly distributed random number from 0 to 1

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

the complement of a random number is a random number

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

taking the natural log

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

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

For constant λ

$$\lambda = 1/MTBF$$

therefore

$$t = -MTBF*(\ln(RN)) \qquad (5)$$

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

$$DBE = -MDBE*(\ln(RN)) \qquad (6)$$

Before the null was rejected, MDBE was the mean for the worldwide fleet. It was a single-valued parameter for a population of known mean. Now MDBE in equation 6 is the mean for a population – the 737 MAX fleet – with an unknown mean. After the first crash, we have a sample of one from a population of unknown mean. After the second crash, we have a sample of two from a population of unknown mean. Small sample sizes suggest there will be great uncertainty, but this should not suggest great inaccuracy.

We now describe developing the departures to next event probability distribution after the second crash as seen in figure 2. This is the probability distribution for departures to a third event. Equation 6 is used as described graphically in figure 1

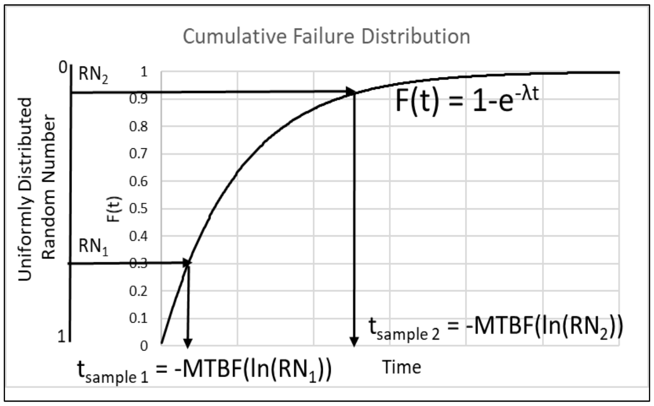

Figure 1- A failure time sample is determined with a random number draw that is transformed to a failure time by the failure distribution. The graph shows two samples of failure time.

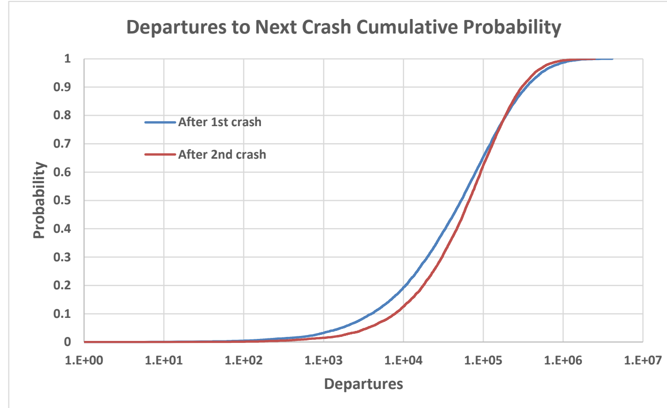

Figure 2 – Probability distributions for departures to next crash immediately upon the 1st and 2nd event. Probability times consequence (fatalities upon a crash) equals risk measured in expected number of fatalities.

to get two DBE samples, DBE_1 and DBE_2. These are averaged for a sample MDBE for our system with unknown mean. Each set of DBE averages provides a sample of MDBE as a random variable ($MDBE_{RV}$). In the same calculation, this $MDBE_{RV}$ sample is used in equation 6 a second time with a different random number. This produces a different DBE sample that is a sample for the departures to next event. The smooth distributions in figure 2 are formed by repeating this process to obtain 10,000 samples. This distribution, using the Poisson interval time, is the time to failure probability distribution that engineers are more familiar with. From the departures to next event distribution, the probability of a third crash within any number of departures is found.

The previous paragraph describes the method using a dataset of two. The method is identical for any number of events. For one event, the $MDBE_{RV}$ sample is the average of only one DBE sample. For 100 events, each $MDBE_{RV}$ sample is the average of 100 DBE samples. A distribution for $MDBE_{RV}$ will narrow as the number of DBE samples increase. The broad distributions in figure 2 are in part due to the small number of samples that form $MDBE_{RV}$. Systems demanding high reliability will always have small major event sample sizes because failures are intolerable. The small sample size leads to uncertainty (broad distributions), but not to inaccuracy.

In the above process of obtaining 10,000 samples for the figure 2 distributions, Monte Carlo simulation is not conducted as in the typical application where a system model is designed to approach reality closely enough to be useful. Simulation is used here to solve a math problem. Equation 6 is used to generate a probability distribution for $MDBE_{RV}$, and that probability distribution is convolved with a distribution for the natural log of an evenly distributed zero to one random number – the right most part of equation 6. So, figure 2 distributions for departures to next crash should be viewed as math solutions and not as computer simulations in the normal sense. Monte Carlo generated random numbers is an efficient way to perform this convolution of two probability distributions. These distributions accurately reflect uncertainty of outcome. Because the addition and multiplication of probability distributions are by Monte Carlo, the accurate results have imprecision. For example, the second crash occurred after an additional 117,856 departures by calculated estimate. To find the probability of a crash by that number of departures, five simulations of 10,000 iterations produced probability results of 0.6793, 0.6871, 0.6914 and 0.6841.

An airworthiness directive (AD) was issued by the FAA nine days after the first crash. Even if the AD totally resolved the problem, the 9-day delay incurred significant risk that is currently unrecognized. There were 227 planes in service during the nine days with an average of 2.85 departures per day, as calculated earlier, resulting in about 5,823 departures in the period. Using data that formed figure 2, the probability of a crash within nine days is about 0.133. The average number of fatalities can be considered the consequence of a crash, or 346/2 = 173 fatalities/crash. Risk is probability times consequence or 0.133 * 173 = 23.0 fatalities on average.

Similarly, following the second crash there was a 3-day delay in grounding the fleet. During the 3-day period there were 384 planes in-service. (Some planes were grounded earlier but Boeing and FAA risk decisions presumably would have been using the entire fleet). The number of departures in the 3-days are 384 * 2.85 * 3 = 3,283 departures in the interval. Using the "After 2nd crash" probability distribution in figure 2, the probability of a third crash is about 0.0459. Risk is 0.0459 * 346/2 = 7.9 fatalities on average. This risk obviously would not have been deliberately taken. There is general lack of knowledge regarding EIP data analysis and especially the long left-hand tail of the unreliability distribution associated with exceedingly small datasets.

6 BOEING 737 MAX PRECURSOR

To avoid the first crash, precursors must be used. Precursors are less severe events that can signal a system problem. The 737 MAX publicly available raw data demonstrate how precursor events and their analysis can potentially avoid even the first crash. Serendipitously and unrelated to precursors, there were data included in the accident investigation report into the October 29, 2018, Lion Air crash [6] that we analyze to identify a precursor capable of avoiding the first crash. Event count data are extracted from the report and are in the left-hand section of table 3 in grey. The green part of table 3 is in-service plane days within the 18-year period from which event counts came, calculated from plane delivery date data downloaded from the Boeing website. Calculations using data in the grey and green region of the table are shown in yellow in the three columns on the right. The single 737 MAX stick shaker activation seen in table 3 may appear to be unimportant relative to other fleet types; however, the relative rate calculation in the right most column puts the 737 MAX into perspective. Relative rates are shown in the figure 3 graph. The 737 MAX is 31 times the Boeing average. The Poisson p-value

Boeing stick shaker activations from 2001 thru 2018 from page 169 Lion Air 10/29/2018 crash investigation report in two left columns in grey, plane-days calculated from delivery dates on Boeing web site in green, yellow columns are calculations. (All Boeing plane-days exclude business and military).

Boeing type	Stick shaker activations	Plane-days in interval	MDBE (plane-days)	Rate	Relative rate
737 MAX	1	64,970	64,970	1.54E-05	30.73
737-300	4	5,758,881	1,439,720	6.95E-07	1.39
737-700/700C/700W	4	5,879,761	1,469,940	6.80E-07	1.36
737-800/800A	18	14,313,713	795,206	1.26E-06	2.51
757-200/200M/200PF	1	5,617,794	5,617,794	1.78E-07	0.36
767-200/200ER	1	969,491	969,491	1.03E-06	2.06
767-300/300ER/300F	1	4,315,924	4,315,924	2.32E-07	0.46
All Boeing	30	59,894,917	1,996,497	5.01E-07	1.00

Table 3 - Stick shaker activations for Boeing fleets from the Indonesian report on the 2018 accident are in the grey two left columns. Plane-days in the interval are calculated from fleet delivery dates from the Boeing website in green. The "stick shaker activations" calculations for MDBE, rate and relative rate are in yellow on the right. Calculations should be obvious.

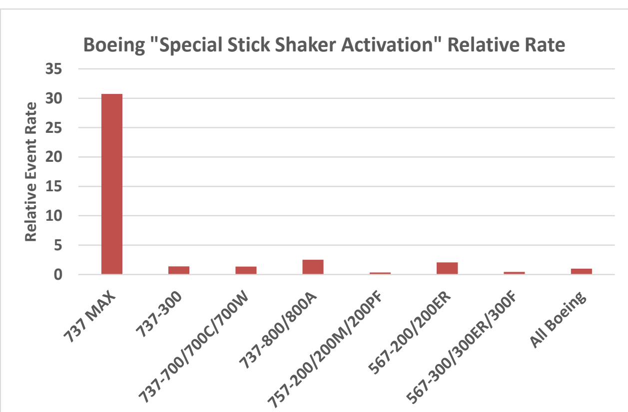

Figure 3 - Stick shaker activation relative rate for all Boeing fleets from table 1. The single stick shaker activation the day prior to the first crash is an obvious outlier – 31 times the Boeing average. The precursor p-value of 0.0235 (0.0136 with prior probability adjustment) is sufficient to reject the null hypothesis that the quick event is due to random variation. The precursor p-value could trigger immediate grounding of the affected plane with investigation.

calculation using the complement of equation 1 is:

$$p\text{-}v1 = 1 - (e^{-\mu})(\mu^{x})/x! = \underline{0.02350}$$

where:

$x=0$

μ = PdBE/MPdBE = 0.02378

PdBE = plane-days between event = 47,481

MPdBE = mean plane-days between events = 1,996,497

The prior probability developed earlier is applicable, so adjusting for the prior, p-v1 = 0.0235 * 0.579 = $\underline{0.0136}$.

With or without prior probability consideration, the precursor could trigger grounding the problem plane for investigation; thereby providing an opportunity to identify any safety issues and their causes. (The data reported in the Indonesian accident report must be a subset of all stick shaker activations. The Australian Transport Safety Bureau [7] reported many more activations on Boeing fleets during a subset of the time, aircraft and geography considered in the Indonesian report. From the context of the crash report, the reported stick shaker activations are most likely those that extended for a long, but unspecified, duration.)

7 BEYOND BOEING 737 MAX

There is nothing unique to the Boeing 737 MAX datasets that render them favorably to the analysis method. Several aircraft designs placed into service with design flaws that were identified and corrected in-service have been analyzed with contemporaneous data [2, 8]. Table 4 is a summary. EIP failed to identify the first crash of the DC 6 in 1947 as being evidence of an unreliable fleet. All others were properly identified as being unreliable upon the first event with no case of a false positive. The FAA (and their predecessor organizations) response record is equivalent to five false negatives.

EIP application is not limited to commercial aviation or engineered systems. In 2020, the death of George Floyd by police officer Derek Chauvin in Minneapolis led to social unrest with protests and riots. EIP analysis of sustained use of force events reveals the former police officer to be an outlier relative to his peer group long prior to the George Floyd incident. Data

Aircraft Type	Design Flaws (discovered early in service)	Fatalities?	Unreliability Identified Upon 1st Event?	
			EIP	FAA and predecessors response sufficient to prevent next event?
DC 6	Fire	yes	no - false negative	no
DC 6	CO_2 in cockpit	yes	yes - true positive	no
DC 10	Cargo door	yes	yes - true positive	no
Concorde	Fire	yes	yes - true positive	not applicable
Boeing 787	Battery thermal	no	yes - true positive(1)	no(1)
737 MAX	Flight control	yes	yes - true positive	no

Note: 1 - Battery thermal runaways while plane is out of service are treated as major events on the assumption these events would have been experienced in the air, but for random

Table 4 - The FAA and predecessor organizations are slow to respond to the first opportunity to recognize and act upon unreliability. The EIP decision record is one false negative, no false positive and five true positives.

analysis results could have affected police complaint investigations, especially with the low false positive probability of 0.001. Incorporating this additional data can improve these investigations. These examples demonstrate what is thought to be universal application of EIP to any event(s).

8 RECOMMENDATIONS

8-1 General

Events can be either desirable or undesirable and intervals can be outliers by being either too long or too short. While EIP can be applied in all four combinations, these recommendations are limited to engineered systems for which events are undesirable, such as for failures (or accidents) of the system.

Rare major events on important systems can simply be analyzed manually by policy and procedure immediately upon system failure. When events are numerous or systems are numerous, the analysis will generally require automation. When even a single event is unacceptable, precursor events should be analyzed, and these will usually require automation of the data analysis process.

Probability values that reject an appropriate null hypothesis are of little value until failure investigation, failure analysis or root cause analysis both reveals and corrects the causes of reliability degradation. EIP simply provides the alarming probability values immediately upon an event, and these trigger the investigation into cause so that future failures can be avoided.

When the null is rejected, the risk of continued uncorrected operation should be assessed. This provides information for risk-based decisions regarding the allowable time to correct. It should be recognized that when intervals are short and the failure dataset is small, the long-left tail of the unreliability distribution extends toward zero. Any uncorrected continued operation of the system may involve substantial risk.

P-values should be viewed as strength of evidence and not necessarily as a “go/ no-go” gage. P-value magnitudes can be used in conjunction with the cost of investigation and the cost of failure. For example, lower cost of failure and higher investigative cost can require lower triggering p-values.

8-2 Specific

The academic community should incorporate EIP into engineering curriculum.

The FAA should incorporate EIP in decision-making following major events and precursors. The NTSB should include the analysis in their event investigations. The industry should recognize both the high probability of any new design being found unacceptable in service and the risk of operating unreliable fleets due to the long-left tail of fleet unreliability distributions. Continuous improvement in the design and certification process has not been sufficient in the past 75 years and should not be 100% relied upon in the future. EIP should be considered a second independent layer of protection to minimize the risk of unreliable aircraft being in service. Such a position by the FAA will be a driving force for manufactures to seek out and learn from any precursors to avoid the potential of a first major event. At some point, attempts to assure the design and certification process is flawless will harm the industry’s commercial viability and prevent development of new and safe products. EIP can help avoid any overreaction to the 737 MAX disasters in trying to make the design and certification process perfect.

REFERENCES

1. J. B. Smith, “Failure Time Analysis Applied to Boeing 737 MAX”, Proc. Ann. Reliability & Maintainability Symp., Jan 2020.
2. J. Smith, “Boeing 737MAX Thru DC6 Fleet Grounding Decisions Revisited with Event Interval Probability Analysis”, Proc. Ann. Reliability & Maintainability Symp. May 2021.
3. J. Smith, “DC 6 Thru 737 MAX – Identifying Unreliability with In-service Precursors to Avoid the First Crash”, Proc. Ann. Reliability & Maintainability Symp. Jan 2022.
4. A. Tangel, A. Pasztor, “Regulators Found High Risk of Emergency After First Boeing MAX Crash”, Wall Street Journal, Aug. 1, 2019 (with link to FAA confidential draft document. The link has been removed, but the document data is visible in the documentary “Downfall: the case against Boeing”).
5. Boeing, “Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations”, 2018
6. Komite Nasional Keselamatan Transportasi (KNKT) Aircraft Accident Investigation Report PT. Lion Mentari Airlines Boeing 737-8 (MAX); West Java Republic of Indonesia, 29 October 2018, final report.
7. Australian Transport Safety Bureau, “Stall Warnings in High Capacity Aircraft: The Australian Context”, Nov. 2013
8. J. Smith, “Failure Interval Probabilistic Analysis for Risk-based Decisions - Concorde Crash Example”, Journal of System Safety Vol. 56 No. 3 Spring 2021.

BIOGRAPHIES

Jan B. Smith, PE
PMF Series, LLC
3342 Courtland Manor Ln.
Kingwood, Texas, USA

e-mail: jansmith@pmfseries.com

Jan Smith, PE, BSME, reliability engineering consultant with a 57-year career in reliability engineering, established two and managed three reliability departments for major corporations. Additionally, he established two professional engineering firms specializing in reliability engineering, for which he was the responsible engineer. He has developed and taught seminars on root cause analysis and supporting engineering methods such as finite element modeling. Most recently, he has developed new probabilistic methods and analysis tools for system availability (PMF Series) and reliability (Event Interval Probability).

Website: https://www.pmfseries.com.