DC 6 Thru 737 MAX – Identifying Unreliability with In-service Precursors to Avoid the First Crash

Jan B. Smith, PE, PMF Series

Precursors, Event Intervals, LaPlace Trend Test, Design Flaws, Probability, Monte Carlo, Event Interval Probability Analysis

SUMMARY & CONCLUSIONS

Before the first design flaw related crash or emergency landing in an aircraft fleet, there are less severe events that signal a reliability problem. The significance of these precursor events is easily recognized in hindsight during crash investigation; however, they are often unnoticed or dismissed prior to a crash or emergency landing. Intentional use of precursors provides an early opportunity to recognize a reliability issue and to eliminate even the first design related crash or emergency landing. This paper reviews the DC 6, DC 8, DC 10, Boeing 787, and Boeing 737 MAX. These planes had design flaws that existed on the day they were placed into service, and they caused major events - crashes and emergency landings. These major events were preceded by precursor events that occurred after the fleet was placed into service, but before the first major event. Of the 937 fatalities caused by the design flaws in this review, 96% are shown to have been avoidable with attention to in-service precursors and data analysis of the precursor and major event intervals.

The precursor analysis uses a newly developed data analysis method - event interval probability analysis (EIPA). EIPA was very recently used on major aircraft events, using data existing at the time of the events. Hundreds of fatalities were shown to be avoidable with incorporation of p-values in fleet grounding decisions [Ref 1, 2, 3]. EIPA applies statistics, probability and reliability theory, and Monte Carlo simulation to event intervals. This paper is focused on application of the method to precursors.

The design flaw, major events and precursors are described for each airplane type. Fleet unreliability is determined with EIPA probability values (p-values). P-values test a null hypothesis that event intervals are consistent with the then existing world-wide jet carrier fleet fatal accident rate. With the null hypothesis that the fleets are as reliable as their contemporaries rejected, computer simulation generates departures to event probability distributions. From these probability distributions, the risk of operation over any future number of departures is determined. This quantifies the unreliability of the fleets and informs as to how fast the design flaw must be corrected and whether the fleet should be grounded.

In the fleet reviews, the early unreliability of the new fleets was unrecognized. Precursors were not sought out by

the industry. When precursors were uncovered, they often were dismissed out of hand. In this review there was no case in which corrective action upon precursors was sufficient to avoid the major event. Meaningful action was taken only after at least one crash or emergency landing. This paper makes the case that precursors should be vigorously sought out and used. The DC 10 and 737 MAX case studies are reviewed in some detail to show why this will require both a paradigm shift in how precursors are viewed and routine data collection with a defined taxonomy and automated contemporaneous event interval probability analysis applied to the in-service precursors. Because statistically significant precursors are rare, there will be a massive number of negative results; therefore, practical application will require management by exception. Routine and automated data analysis should screen and report for review only the very small number of precursor interval p-values that flag need for investigation.

This paper has nothing to do with design and development of the aircraft. It considers events that occur after the plane is placed into service. This allows automated EIPA to monitor and alarm on in-service aircraft fleet reliability/safety performance problems independently of the design and certification process. While assurance of the safety of new designs is being continuously improved, growing design complexity will continue to assure lack of perfection. Both history and common sense suggest that independent inservice performance monitoring is required.

1 INTRODUCTION

In general, the commercial aviation industry has an admirable safety record with amazing improvement over time; nonetheless, there is a small subset of lack of knowledge that is resulting in bad decisions, easily avoidable fatalities, and unnecessary negative business impacts. Fleet grounding decisions based upon Event Interval Probability Analysis (EIPA) p-values, using only data that existed at the time of the major event, has been shown to significantly outperform the historical fleet grounding decisions over the past 75 years [Ref 1, 2]. Application to the Boeing 737 MAX, again using only data existing at the time of the crashes, demonstrates the easily avoidable 737 MAX disastrous circumstances with the EIPA application of statistics, probability and reliability theory, and Monte Carlo simulation [Ref 3]. However, the industry, as well as academia, have not yet demonstrated acceptance and

use of this data analysis methodology. It was first published in 2018 [Ref 4], so the lack of acceptance and use of this new data analysis method thus far is somewhat understandable.

Obviously, there is lack of awareness of EIPA and the applicability to product safety when it is not applied even to major disasters. So why bother with precursors when we are not yet applying the data analysis method to even major event intervals? Hundreds of fatalities could have been avoided with EIPA application to fleet grounding decisions, but hundreds more could have been avoided with application to precursors to avoid even the first major event. So, application to precursors must be communicated.

Precursors are relatively minor events that announce major events that are likely to occur in the future without action. These are easy to identify in hindsight after the major event occurs and they aid in understanding the major event, but precursors are difficult to identify before the major event. Minor events that could be a precursor to a major event are often discounted because we believe our design to be reliable and we resist any evidence to the contrary. Furthermore, nearly all minor events are not precursors, but only when interval p-values so suggest. To proactively use precursors to avoid the major event, precursors should be intentionally sought out and, where appropriate, EIPA applied to identify the earliest opportunity to intervene to avoid future events that otherwise are to be expected. Separation of the signal from the noise, the few true precursors from the numerous minor events of no consequence is demonstrated with the 737 MAX and DC 10, that together led to 692 avoidable fatalities.

EIPA results for the Boeing 737 MAX major events are reviewed to demonstrate both the method and the present absence of consideration of event intervals in decisionmaking, even regarding multiple fatal crashes.

Precursors applied to the 737 MAX example shows how a single precursor p-value was sufficiently strong to avoid the first major event. It also demonstrates how even serious precursors that should be obvious even without EIPA are often unrecognized.

The DC 10 example demonstrates how precursor events can individually be insignificant. Yet though tedious and usually insignificant, these must be immediately analyzed to allow immediate action when probability alarms so indicate. This is exactly what humans are not designed to do, but computers are. Automated data analysis will be necessary for implementation on precursors. Grouped data is demonstrated for single datasets with the DC 10 cargo door problems. Such reported data may be grouped by month with departures by aircraft and aircraft type identically grouped. Automated data analysis for 15,000 assets and every work order has been done with Microsoft Excel [Ref 4]. This will most likely be an inadequate big data tool for the proposed analysis, but it places the analysis problem into perspective – automated analysis in a big data manner is practical and demonstratable.

All examples in this paper underscore important but unappreciated facts, 1) all these fleets were placed into service while they were unreliable on day one, 2) the industry speaks and acts as though they are unaware of the above, so we presume they are unaware.

2 BOEING 737 MAX

EIPA has been retroactively applied to the 737 MAX departure intervals between crashes and reported [Ref 3]. The probability of the first crash occurring that early in the fleet life, when compared with its contemporaries, is 0.022. This low probability should reject a null hypothesis inherent in EIPA that the failure data are generated by a homogeneous Poisson process with events being independent and identically distributed exponential random variables. The alternative hypothesis should be accepted that the fleet is less reliable than the worldwide commercial jet scheduled carrier fleet with statistical significance. The number of departures to the first crash signals a likely problem with the fleet.

The probability of the two consecutive crashes within the number of fleet departures at the time of the second crash is 0.00099. This low probability overwhelmingly rejects the null hypothesis. But EIPA was not applied. The methodology was only first reported in 2018 and is relatively unknown. The probability of a third crash during a 3-day delay in grounding, assuming all planes are flying, is found to be 0.0475 with a risk of 8.22 fatalities. Of course, this significant risk was unrecognized and was avoided only by chance.

Figure 1 is three cumulative probability distributions that captures the essence of EIPA results for the 737 MAX. These distributions are obtained from the cumulative failure distribution function, equation 1, starting with the familiar units of time (t) and failure rate (λ) and mean time between failure (MTBF). Then we change to the more usable and equivalent units of departures between events (DBE) and mean departures between events (MDBE) as follows:

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

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

$$\lambda = 1/MTBF$$

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

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

$$DBE = -MDBE*(ln(RN))$$
(3)

MDBE in equation 3 is the mean for the 737 MAX contemporaries – the worldwide mean.



Figure 1

The middle distribution is DBE for the worldwide fleet. The 1st crash probability is 0.022 from this distribution, warning of low reliability. The right distribution shows about 0.001 probability of occurrence of the two quick events – much too unlikely to be random chance. Rejecting the null, the actual DBE distribution after two crashes is on the left and shows high risk of a third crash within the number of departures expected in the 3-day grounding delay.

Boeing Type	"Stick Shaker Activations" 2001 - 2018	Plane-days in interval	MTBE	rate	relative rate
737-300	4	3593699	898425	1.113E-06	2.285
737-700/700C/700W	4	5893478	1473370	6.787E-07	1.393
737-800/800ER	18	14381019	798946	1.252E-06	2.569
737-MAX	1	70,732	70732	1.414E-05	29.023
757-200/200M/200PF	1	3562424	3562424	2.807E-07	0.576
767-200/200ER	1	819718	819718	1.220E-06	2.504
767-300/300ER/300F	1	3013327	3013327	3.319E-07	0.681
All Boeing	30	61584737	2052825	4.871E-07	1.000

Table 1

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. The "stick shaker activations" calculations for MTBE, rate and relative rate are in yellow on the right. Calculations should be obvious.

Although even the primary events were not analyzed, we now demonstrate how precursors, and their analysis can potentially avoid even the first crash. The accident investigation report into the October 29, 2018, Lion Air crash [Ref 5] included the left-hand section of table 1 in grey. This table is extended to reveal the in-service precursor. The green portion of table 1 is in-service plane days within the 18-year period as reported in the accident report, calculated from



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 – 29 times the Boeing average. The precursor p-value of 0.0339 is sufficient to reject the null hypothesis that the quick event is due to random variation, especially given it is a new design. The precursor p-value could trigger awareness.

delivery date data downloaded from the Boeing website. At first glance the single 737 MAX stick shaker activation is unimportant relative to other fleets. But the relative rate calculation in the right most column puts the 737 MAX into perspective. Relative rates are plotted in figure 2. The 737 MAX is 29 times the Boeing average. The Poisson p-value calculation (as in references 1-4) is 0.0339 using data from table 1. This p-value is the probability of the 737 MAX stick shaker activation interval being that short relative to the Boeing average by random chance. (The data reported in the Indonesian accident report must be a subset of all stick shaker activations. The Australian Transport Safety Bureau [Ref 6] 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 only those that extended for a long, but unspecified, duration.)

With Bayesian reasoning, we can add to our information that new fleet designs are frequently placed into service with design flaws that cause crashes or emergency landings. Considering only Boeing major types and excluding the MAX, three of the last ten have had a design flaw leading to groundings of all or part of the fleet - 777 P&W engine blade failure, 787 battery fire, 737 rudder control. The equivalent of the Bayesian prior is 70% chance of being reliable day one. Combining these independent probabilities, 0.0339*0.7 =0.0237. With or without this p-value adjustment, we see that the stick shaker activation precursor p-value was capable of triggering investigative action.

But the rarity of the 737 MAX precursor is much greater than the low p-value suggests. The precursor experienced prior to the first crash was a loss of control of the aircraft that is both much more severe and rare than a stick shaker activation. Extended loss of control should have been treated as though it were a crash. This is easy to say after a major event, but in the future we should and can do so before the major event. Even though precursors are easy to identify after the fact, it was not even mentioned in the Indonesian investigation report [Ref 5] in the context of an overlooked precursor capable of having avoided the first crash. This indicates that the industry currently does not view precursors as important.

3 DC 10

On June 12, 1972, a DC 10 cargo door latch system failed in flight with explosive decompression blowing a hole in the hull and partially collapsing the passenger floor onto the control cables. The wrecked plane was landed without fatalities, but there was significant risk. The manufacturer placed responsibility on the ground crew responsible for closing the door; nonetheless, some door latch improvements were initiated [Ref 7]. The issue was not recognized and treated as a serious design fault. Using EIPA, it has been shown that grounding of the fleet with assurance of design correction would have been appropriate with only 0.0013 probability of false positive [Ref 1, 2]. While even the significance of the timing of the near fatal crash was ignored, how could consideration of precursor events have allowed this near fatal accident to have been avoided? How could such accidents be prevented in the future?

It is reported that there were "approximately 100" prior issues with the cargo door throughout the fleet since being placed into service in 1971 [ref 7 page 152]. To demonstrate method, EIPA is retroactively applied to these door issue events and intervals to see if probability values could have drawn attention to a reliability problem. Actual intervals are unavailable and only the approximate total over fleet life at the time of the major event is known. A stochastic model is used to generate exemplar door issue time between event data as a random variable. These in turn allow LaPlace trend test pvalue distributions and confidence intervals.

Figure 3 is the age of the DC 10 fleet at the time of the major event. Age is calculated from delivery dates from the Boeing website. (As downloaded, delivery month and year was available, but not day. All deliveries for the month were treated as if delivered on the 15th.) The specific plane that failed was one of the two oldest aircraft in this new fleet. As described in reference 7, the door latching mechanism failed due to use and adjustments. There is no indication that a wear-out failure mode on the oldest plane drew any attention by the manufacturer, the FAA or NTSB, much less that the probability of such an event so early in the fleet life approached zero [Ref 1, 2]. The manufacturer would argue then, and perhaps even today after the 737 MAX lessons learned, that such in-service problems are not within the scope of the design. But design features that facilitate failures due to



Figure 3





Figure 4



imperfect operation and maintenance should be within the design scope. (Whether the door issues are accepted as a design fault or not, the manufacturer must later correct the design, but only after 346 fatalities. Furthermore, the findings of the inevitable congressional investigation that follows a major loss of life event could nearly have been copy/pasted from that of the future 737 MAX or prior investigations back to the DC 6 in 1947).

Figure 4 shows the concept for generating door issue data to replicate the unavailable real data. This is only to demonstrate EIPA process and capability - not to get an absolute answer. A Weibull shape parameter of 2.5 and location parameter of 285 generates a model fleet mean of 99.5 with 90% confidence interval of 76 to 126. (This spreadsheet model with video description is found at reference 8). Event intervals and simulated event counts as random



Table 2

Interval and counts within the interval for each iteration is input data in spreadsheet columns AI and AJ in green. Cell formulas are shown with LaPlace trend test p-values in the right most column. P-value probability distributions are formed with numerous iterations. The model and description are available [Ref 8].

variables are in the green columns in table 2, which in turn are used to calculate LaPlace trend test p-value random variables in the right most column. The LaPlace trend test [Ref 9] is typically applied to single events. In table 2, variable event intervals and counts within those intervals, including a count of zero, are grouped to allow the LaPlace trend test to be used. P-values show degradation of the cargo door over time. Cell AQ14 is the p-value after 333 days in service. The mean pvalue is 0.02613 with 90% confidence interval of 0.0014 to 0.0899. The null hypothesis of no trend can be rejected. Cell AQ12 is 62 days earlier with mean p-value of 0.0479 and 90% confidence interval of 0.0039 to 0.1561. It is likely that degradation this early could be detected, allowing intervention to stop the failure mode. This is especially true considering that some planes obviously did degrade and will degrade faster than the fleet average that the model generates.

In practice, precursor events are generated by the system behavior – not calculated. The model-generated events in figure 4 and table 2 are to demonstrate analysis method. The table illustrates how events may be gathered over an interval such as monthly for both LaPlace and Poisson p-values and pvalue trends. For implementation of the concept, precursor events would be gathered from safety reports or work orders received electronically and the analysis would be automated. When there are statistically significant precursor events flagged by low probability values, an exception report would be generated by the automated system to bring awareness of the possible problem for investigation.

On March 3, 1974, another cargo door opened in flight with explosive decompression and 346 fatalities. The latching

mechanism improvements made since the 1972 near fatal crash were via a routine service bulletin and not fully implemented [Ref 7 page 157]. There were 550 failure mode related issues reported with the cargo door for the fleet in the six months before this crash. This excludes door sealing issues not directly related to door latching [Ref 7 page 250]. The number of in-service days for the DC 10 fleet were calculated from delivery dates from the Boeing website in the same manner as before. There were about 21,819 DC 10 aircraft operating days in this 6-month period. The 550 door issues produce a door problem rate of 0.0252 problems per plane-day. This average problem rate is about the same as at the time of the first wreck, which is approximately 100 problems divided by 4,415 fleet plane-days, or 0.0227 problems per plane-day. This lack of improvement in the door issues after the first nearly fatal accident underscores the underreaction and ineffectiveness of any door latching system improvement made earlier. More generally, it reflects a lack of awareness that both major and precursor event probabilities exist and are calculable. Instead of treating early system failures as an unfortunate event on a reliable fleet, we should aggressively look for indication of any system weakness.

4 OTHER DESIGN FLAWS & AIRCRAFT TYPES

The DC 6 was grounded in 1947 because of fuel spillage during inflight transfer with fires. There are no known precursors, but this may well be because it is old history and evidence is missing. The design improvement included fire suppression. This resulted in CO_2 in the cockpit with loss of life preceded by resistance to a warning precursor [Ref 7]. The DC 8 had human factors design problem in the cockpit not recognized until in service for 10 years [Ref 7]. The Boeing 787 had a battery fire problem. The precursor (treated as a major event in prior papers) was an on-ground battery fire. A few days later a fire forced an emergency landing, and the fleet was grounded. Like the DC 6, when the issue results in a fire, there is little to blame but the design, so acceptance that the design has an unacceptable flaw comes quicker.

All designs and flaws are summarized in table 4. Of the 937 fatalities, 96% could potentially have been avoided with attention to both precursors and major events and data analysis of the event intervals. This summarizing table perhaps oversimplifies complex situations and may well be unfair to organizations and individuals, but it underscores a valid point - attention to in-service major events and precursors can save lives and avoid negative business consequences such as in the 737 MAX case.

5 RECOMMENDATIONS

1- The industry should incorporate in-service precursors and their intervals, along with major event intervals, to monitor and assess the reliability and safety of in-service aircraft and aircraft fleets.

			1				1
Plane				In-service	Initial Industry Response to	Initial Industry Response to	Primary Event
Type	Design Flaw	Primary Event	Fatalities	Precursors?	Primary Event	Precursors	Avoidable?
	In flight fuel spillage				· · ·		
DC 6	w/fire	1947 crash	40	none known	no known response	no known precursors	no
	In flight fuel spillage	emergency landing in					
DC 6	w/fire	flames	0	none known	grounded DC 6 fleet	no known precursors	no
	CO ₂ concentration in						
DC 6	cabin	1948 crash	35	yes	grounded DC 6 fleet	resisted precursor	yes
	Cockpit human factors						
DC 8	design	1970 crash	109	see crash history	none	none	likely yes
	Cockpit human factors						
DC 8	design	1971 crash landing	0	yes (1971 crash)	none	none	yes
	Cockpit human factors						
DC 8	design	1972 crash	61	yes (prior crashes)	begin study, slow implementation	begin study, slow implementation	yes
	Cargo door latching	In-flight failure			ground crew faulted, minimal	door issues p-value trend not	
DC 10	system	w/emergency landing	0	yes	change, slow implementation	recognized	yes
	Modified cargo door				ground crew faulted, redesigned,	prior failure and precursor p-	
DC 10	latching system	1974 crash	346	yes	fleet not grounded	values not recognized	yes
Boeing 787	Battery fire	emergency landing	0	yes	grounded fleet	start investigation	yes
							possibly yes /
					Relied upon the risk assessment	loss of flight control unrecognized	with automated
737 MAX	Flight control	2018 crash	189	yes	methodology that allowed the flaw	as precursor	analysis - yes
				yes (2018			
				predecessor and	Relied upon the risk assessment	Probability of 2 quick crashes	
737 MAX	Flight control	2019 crash	157	crash)	methodology that allowed the flaw	unrecognized	yes

Table4

This table summarizes 7 design flaws (considering the DC 10 cargo door as 2 different designs) on 5 aircraft types. All but the first design flaw on the DC 6 in 1947 had identified precursors. Of the 937 fatalities, 96% could potentially have been avoided with attention to precursors and primary events and data analysis of event intervals.

2- Air worthiness certification should include the above as a continuing requirement and the FAA should facilitate methodology.

3- EIPA is new and has not yet been incorporated into the reliability engineering body of knowledge. The engineering academic community should consider addressing this situation.

4- The general applicability of EIPA to engineered system failures and precursors should be recognized by other industries. Applicability to non-engineered and abstract systems should be recognized by other professions.

REFERENCES

- Smith, J. B., "Failure Interval Probabilistic Analysis for Risk-based Decisions – Concorde Crash Example", *Journal of System Safety*, Volume 56, No. 3, International System Safety Society, Spring 2021.
- 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. B. Smith, "Failure Time Analysis Applied to Boeing 737 MAX", Proc. Ann. Reliability & Maintainability Symp., Jan. 2020.
- J. Smith, O. Sac, K. Bordelon, "Contemporaneous Failure Time Analysis Using Poisson Probability", Proc. Ann. Reliability & Maintain Symp., (Jan.) 2018.
- 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.

- Australian Transport Safety Bureau, "Stall Warnings in High Capacity Aircraft: The Australian Context", Nov. 2013
- 7. P. Eddy, E Potter, B. Page, *Destination Disaster*, New York, Times Newspapers, Ltd., 1976.
- 8. Stochastic model is at <u>www.pmfseries.com</u> for download with link to explanatory video.
- 9. P. O'Conner, A. Kleyner, *Practical Reliability Engineering*, fifty edition, John Wiley & Sons, Ltd, 2012.

ACKNOWLEDGEMENTS

Three London Times reporters, Paul Eddy, Elaine Potter, and Bruce Page did the tedious research reported in the 1976 reference 7 that qualitatively describes missed opportunities to avoid crashes and emergency landings that we call precursors.

BIOGRAPHIES

Jan B. Smith, PE PMF Series, LLC 3342 Courtland Manor Ln. Kingwood, Texas 77339 USA

e-mail: jansmith@pmfseries.com

Jan Smith is a Reliability Engineering Consultant with a 57 year career in reliability engineering, principally in repairable systems. He has established and managed reliability departments and professional engineering firms as well as developed and taught root cause analysis client specific and public seminars. Most recent work is developing new probabilistic methods for system availability and capacity (PMF series) and reliability (EIPA), with automated analysis contemporaneous with events. Website: www.pmfseries.com.