Redefining the Incidents to Learn From

Citation for published version:
Walker, GH 2017, 'Redefining the Incidents to Learn From: Safety Science Insights Acquired on the Journey from Black Boxes to Flight Data Monitoring', *Safety Science*, vol. 99, no. Part A, pp. 14-22.
https://doi.org/10.1016/j.ssci.2017.05.010

Digital Object Identifier (DOI):
10.1016/j.ssci.2017.05.010

Published In:
Safety Science

Redefining the Incidents to Learn From: Safety Science Insights Acquired on the Journey from Black Boxes to Flight Data Monitoring.

Dr Guy Walker

Heriot-Watt University, Edinburgh, [UK] EH 14 4AS

ABSTRACT

The reason Flight Data and Cockpit Voice Recorders (FDRs and CVRs) exist is to learn from incidents. Probably no other single invention has yielded such significant improvements in aviation safety. Indeed, they have been so effective that we now need to redefine what is meant by the term ‘incident’ and the uses to which data recording technologies are now put. The paradox is that at no previous point in history have we collected so much data, yet safety performance is such that it is rarely used for its original purpose: as a lagging indicator of problems following an accident. In this paper the history of black boxes is briefly surveyed and connected to the underlying safety science knowledge base. Flight Data Monitoring (FDM) is then presented as an exemplar of the paradigm shift from lagging to leading indicators needed in order to continue learning from incidents. In many industries the pre-requisites for comparable Data Monitoring processes are already in place. The benefits to be accrued by following the example set by the aviation industry are considerable.

Keywords

Leading and lagging indicators, data recording, flight data monitoring, safety management systems, risk triangle

INTRODUCTION

Where are the Incidents?

2010 represented a landmark year in European aviation safety. This was the first year in the entire history of European aviation that no fatal commercial air transport incidents occurred at all (EASA, 2010). Whether it takes into account exposure by distance (0.01 fatalities per 100 million miles) or number of flights (3.1 fatalities per 10 million) the risk to the travelling public within the European aviation sector is exceedingly low (EASA, 2010). In everyday terms the probability of a fatal air crash in European airspace is approximately equal to winning second prize on the Euro Lottery (a jackpot of over €0.5 million). The highest risk for most travellers stems from the car journey to the airport (approx. 1 in 20,000 chance of a fatal accident; WHO, 2016) and using the escalators once inside the terminal building (65% of all such accidents occur in transport facilities; Schminke et al., 2013).

This safety science success story has, to a very considerable degree, benefitted from a device the sole purpose of which, was to help the industry learn from incidents. This device is the humble ‘black box’, or to put it more correctly, the combined features of a Cockpit Voice Recorder (CVR) and Flight Data Recorder (FDR). This paper takes the opportunity of a Special Issue on Learning from Incidents to present the brief history of black boxes and the often unstated relationships it has with fundamental safety science principles. It also reveals a strong paradox instructive for other domains. Never before has there been access to so much data, and the ability to learn from incidents, yet serious fatal incidents from which to learn have (in the aviation sector at least) almost disappeared. Powerful trends in the safety science domain are discussed, including the role of lagging, coincident and leading indicators and a new role for black boxes described. Instead of a post-hoc accident analysis tool, something that needs an accident as a ‘lagging indicator’ of future risks, black boxes can be used as a predictive safety assurance tool, a supplier of ‘big data’ from which leading indicators of strategic risks can be derived in novel ways.

BRIEF HISTORY OF THE BLACK BOX

Origins

The act of automatically recording data on system parameters over time is referred to as "data logging" or "data recording". In the aviation industry it falls under the specific heading of Flight Data Recording, which comprises several individual procedures and devices. The most prominent device is colloquially termed a 'black box'. This represents the combination of a Flight Data Recorder (FDR) and a Cockpit Voice Recorder (CVR). Other systems under the heading of Flight Data Monitoring include various Aircraft Condition Monitoring Systems (ACMS), such as engine health monitoring (e.g. the Rolls Royce EHM programme) and the wide range of parameters available from modern avionics (e.g. ARINC 573) via so-called 'Quick Access Recorders' (QARs).

Data logging can trace its formal origins to the allied fields of metrology, instrumentation, telemetry, predictive maintenance and functional performance. (Campbell, 2007). The use of data logging as a tool in post-accident analysis, and safety more generally, is a comparatively recent development. It was led by the invention of the original 'black box' by Dr David Warren working for what would become the Australian Defence Science and Technology Organisation (DSTO) in the early to mid-1950s.

The DeHavilland Comet crashes of 1953 and 1954 provided particular motivation for the use of FDRs. The DeHavilland Comet was the world's first commercial jet airliner. It revealed itself to have problems with metal fatigue resulting in catastrophic failure of the pressurised fuselage. At the time there was very little data to inform subsequent investigations. Numerous technical committees were instituted to examine the crashes and a report entitled "A Device for Assisting Investigation into Aircraft Accidents" (1954) was produced. It suggested that *"anything which provides a record of flight conditions, pilot reactions, etc. for the few moments preceding the crash is of inestimable value"* (Warren, 1954).

Devices for in-flight condition monitoring did exist at the time. Early examples included the NACA2 V-g recorder. This was a device used in transport and bomber aircraft during World War II to assess operational loads (Campbell, 2007). These so-called 'analogue/analogue' FDR devices relied on film exposed to light traces or styluses leaving a physical impression on rolls of Incanol steel. So called 'scratched foil' technology continued to be used in early FDRs well into the 1950's, appearing on Boeing 707s from 1958 onwards and pre-dating mandatory fitment by a number of years. Despite their safety benefits, these early recorders were not durable, measured only a few parameters, required significant effort to interpret and could not record voice transcripts. Further development was necessary.

Early Black Boxes

The technical capability for a more advanced FDR, one that included voice transcriptions, arose from a piece of consumer audio technology called the Miniphon. Manufactured by a West German electronics company called Protona GmbH, the Miniphon used a fine coil of stainless steel wire that passed from one reel to another over a magnetising “head”.This enabled low quality, but nonetheless intelligible and durable voice recordings to be made. The prototype ‘black box’ (at this point referred to formally as a ‘Flight Memory Unit’) used parts from a Miniphon recorder combined with other electronics that could superimpose signals from some of the aircraft’s primary controls onto approximately 30 feet of metal wire, at a rate of eight signals per second. The device was configured so that the metal wire looped continuously, storing four hours of voice and data and continually over-writing itself. It was then installed into a ‘crash survivable’ enclosure which could be removed from an aircraft and the recordings interpreted in the laboratory.

In 1958 the UK Air Registration Board became aware of the Flight Memory Unit. Due to the national importance of the UK jet aviation industry and the potential safety barrier the Comet crashes represented for continued foreign sales the concept was considered important enough to warrant further development. A clock manufacturer named S. Davall and Sons acquired production rights and developed the first commercial ‘black box’. This became known as the “Red Egg” and is formally called the “Davall Type 1050”. Notable improvements made by Dr Warren and his team now enabled readings to be captured at a rate of 24 per second, and assured greater accuracy in the data collected from aircraft instruments and controls. It also became possible to to record voice or data, or both together. To do this up to 40 miles of stainless steel wire was needed as a recording medium.

Figure 1 – The Davell Type 1050 Flight Data Recorder, or 'Davell Red Egg', so called because the London clock making firm of S. Davall & Sons Ltd manufactured it and its (red coloured) spheroid shape contained an improved magnetic wire recorder. The shape was adopted due to its location in the unpressurised tail sections of early jet airliners, and to give it sufficient strength to be crash survivable. This was the first commercial Flight Data Recorder (Source: Campbell, 2007).

Australia was the first country to make the use of Cockpit Voice Recorders (CVRs) mandatory (DSTO, 2005). The USA was quick to follow with regulations appearing in 1960 making it mandatory to carry a flight data recorder on passenger-carrying aircraft (Morcom, 1970). Similar developments were underway in the UK and changes were made to the Air Navigation Order as early as 1960, although a lengthy period of consultation and evaluation ensued. The first crash investigation, in the UK, to make substantial use of the data provided by an FDR occurred in 1965, the year they became mandatory. This accident involved a BEA Vanguard fitted with a Davell Type 1050 'red egg' which crashed in poor weather at London's Heathrow airport (AIB, 1965).

In the UK the mandatory installation of CVRs took longer. Proposals to install the devices were drawn up in 1969 by the Directorate of Flight Safety and the Department of Trade & Industry, in conjunction with the Air Accidents Investigation Bureau (AAIB), but were met with opposition. A Working Party was formed in 1970 to examine the issues further. Based on this advice the Civil Aviation Authority (CAA) proposed amendments to the Air Navigation Order to make their installation mandatory by 1975. The Staines air crash in 1972, involving a BEA Trident proved decisive. The judge who oversaw the public inquiry was frustrated by the absence of CVRs and concluded the following: *"The investigator is still left in the dark as to what was passing between the crew members by way of orders, comment or exclamation. [..] It seems to us that a requirement for the installation of cockpit voice recorders in airline aircraft (i.e. those over 27,000 kilograms all-up weight) is overdue"* (AIB, 1973, p. 56).

Figure 2 – Cockpit Voice/Data Recorder (CVR/FDR) recovered from the crash site of the German Wings Airbus A320 operating Flight 4U9525 (Source: Reuters).

Development of the Black Box

Early FDRs were relatively stand-alone devices. The recorder carried its own sensors and, apart from an electrical supply, operated relatively independently of the host aircraft (Campbell, 2007). As a result calibration proved to be a problem. The actual state of the aircraft systems was not necessarily identical to those indicated by sensors in the recorder or even sometimes the same as those indicated to the pilots on their cockpit displays. Parallel technical developments in the emerging field of avionics were to see FDRs and CVRs become an increasingly integrated part of the aircraft. Rather than stand-alone devices, data recorders are now part of a comprehensive data acquisition architecture. Data is integrated from a myriad of sources via a Flight Data Acquisition Unit (FDAU), common communications protocols (ARINC 573, 717 and 767) and the use of quick access recorders as well as crash survivable 'black boxes' (Figure 3). The separation between a crash survivable data and voice recorder (mandated by law and used for accident investigation)and a Quick Access Recorder (QAR - not mandated but used for operational and safety purposes by airlines and regulators) occurred in the 1970s. It arose from a growing recognition that easy access to flight data, both routine and abnormal, was of value.

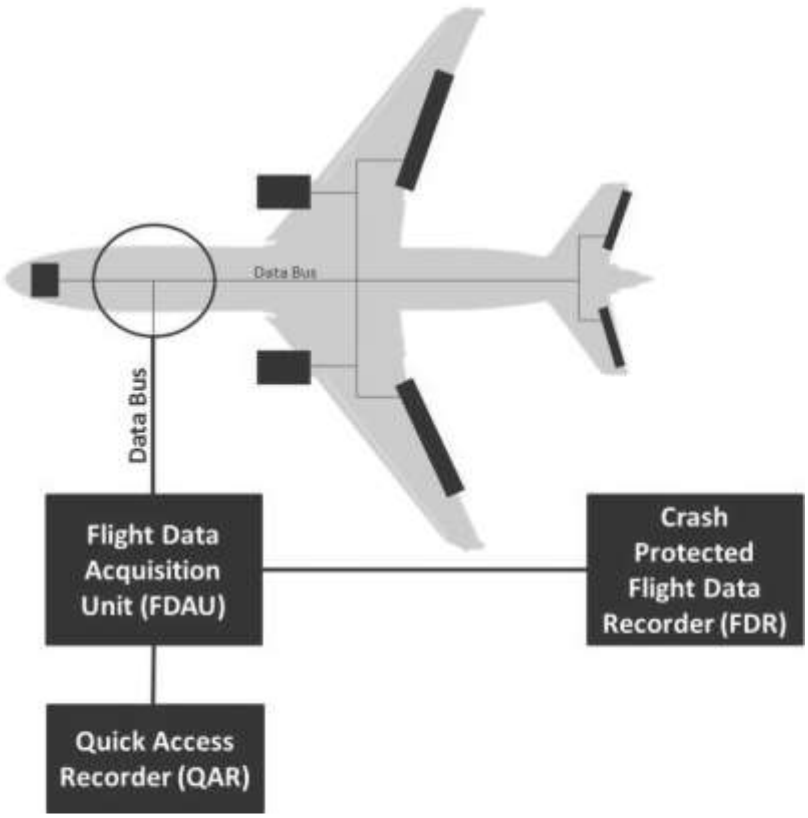

Figure 3 – High level aviation data recording architecture

The use of recorder data for the characterisation and monitoring of routine flight operations had begun in 1964 under the UK's Civil Aircraft Airworthiness Data Recording Programme (CAADRP). A small sample of aircraft had been equipped with paper trace recorders in addition to those fitted by mandate. These were to provide data on aircraft performance in routine operations, such as landings (Hutton, 1972), turbulence (King, 1967), metrological conditions (King, 1967) and so on. This programme ran for over forty years (finishing in 2005) and provided a powerful demonstration of the benefit of data recording as a tool for safety and operational improvements. Simultaneous with the CAADRP programme was the development by Hawker Siddeley and Smiths Aviation of the world's first autoland system. Due to its comparatively advanced avionics the HS 121 Trident aircraft was selected as the prototype. The avionics allowed the autoland system to interface with aircraft systems in a way not possible with more conventional aircraft. It also allowed ready access to data from those same systems via a new device: the Quick Access Recorder (QAR). Data recording enabled autoland to be tested and certified. The Trident became the first commercial aircraft fitted with this system and a QAR. The value of these devices (and the supporting architecture) continues to be demonstrated. As installed today on modern aircraft, QARs record vastly more parameters than the mandated FDR/CVR ones (more than 2000 in some cases) and for a longer time period (more than 30 days). The QAR enables this data to be easily extracted from the aircraft, either by removable cassettes, memory cards or, in some current systems, wirelessly and automatically while the aircraft is parked on stand. As Table 1 shows most modern airliners are able to record thousands of channels of operational flight data; consequently the learning possibilities from incidents have never been greater. The paradox is that serious incidents to learn from are now rare.

Table 1– Development of the black box (Source: Campbell, 2007)

Generation	Aircraft	Date	Data Recorder Type	Parameters Measured	Recorder Capacity
First Generation Foil/Photographic Recorders	Boeing 707	1958	Analogue (foil)	5	Mechanical limit of approx. 5 parameters
Second Generation Wire Recorders	Vickers Vanguard Trident 2	1965 1968	Analogue (wire)	17	300 hours
Third Generation Tape Recorders	Boeing 747-200	1971	Analogue (tape)	32	
	Airbus A300	1974	Digital (tape)	280	128 12bit words per second via serial data input
Fourth Generation Solid State Recorders (SSFDR)	Airbus A321	1988	Digital solid state audio and flight data	450	256 words per second via serial data input
	Boeing 777	1995		1400	
	Airbus A380	2007		2,000	1,024 words per second via serial data input
	Boeing 787	2009	Digital solid state audio, flight data and video	2,200	Ethernet/CANbus systems

LEADING AND LAGGING INDICATORS

The Black Box Paradox

The retrieval of a crash survivable FDR/CVR is a common trope in the popular media but is far from common in reality. The vast majority of airliners will spend/have spent their entire 30 or 40 year lifespans with their black boxes never being used for their primary purpose. This paradox, combined with the growing abundance of data, is the reason for a switch in emphasis from lagging to leading indicators. This is an important fundamental concept for safety science, not often explicitly stated or reviewed, but one that is likely to represent the future paradigm. It drives to the heart of this Special Issue in terms of how we should ‘learn from incidents’, the value of doing so, indeed, what should even be regarded as ‘incidents’ in the first place.

Definitions

The concepts of leading and lagging indicators are well used within the safety science literature (e.g. Shea et al., 2016; Walker & Strathie, 2015; Sinelnikov, Inouye & Kerper, 2015; Hinze, Thurman & Wehle, 2015; Reiman & Pieltkainen, 2012 and others) but it is worth restating precisely what they mean. Leading indicators are measurable precursors to major events such as an accident. The indication of a precursor ‘leads’, or comes before, the actual event itself. Lagging indicators are ‘loss metrics’ that only become apparent after an event (Rogers, Evans & Wright, 2009). Leading indicators are said to be ‘proactive’ because they enable steps to be taken to avoid seriously adverse consequences. Lagging indicators are said to be ‘reactive’ in that a seriously adverse event needs to occur before it can inform learning. For this reason leading indicators are also sometimes referred to as ‘positive performance indicators’ and lagging indicators as ‘negative performance indicators’. Both are sometimes referred to as ‘key performance indicators’, with leading indicators being ‘upstream’; lagging indicators ‘downstream’. Coincident indicators are neither: these are measurable indications that vary at the same time as the trend to which they are related. Table 2 provides some examples of leading and lagging indicators.

Table 2– Examples of leading and lagging indicators

Leading Indicator	Lagging Indicator
Falling barometric pressure is a leading indicator of worsening weather.	A thunderstorm is a lagging indicator of low atmospheric pressure.

Coolant temperature is a leading indicator of problems with an engine.	A failed head gasket is a lagging indicator of a cooling problem.
High cholesterol is a leading indicator of poor coronary health	A heart attack is a lagging indicator of high cholesterol.

It is self-evident that “there must be an association between the inputs that the leading performance indicators are measuring and the desired lagging outputs” (p. 3). While self-evident it is far from formally tested in many cases. It is quite possible for good performance to be achieved despite poor inputs and poor performance to be achieved despite good inputs. Within the oil and gas sector these set of contingencies are presented in a matrix (Figure 4). The horizontal and vertical lines dividing the square into four cells are based on previous baseline performance and the values of chosen leading and lagging indicators are used as coordinates to plot actual safety performance. The advantage of this matrix view is that favourable leading indicators do not simply cancel out undesirable lagging indicators. In other words, simply setting easy targets will move the organisation from Quadrant 1 (poor input and output) to Quadrant 3 (good input but poor output). Likewise, ‘simple good fortune’ will move an organisation to Quadrant 2 (good output but poor input). In fact, the only way for an organisation to move into Quadrant 4 (the ‘continuous improvement zone’) is to set appropriate targets and achieve genuine safety performance against them.

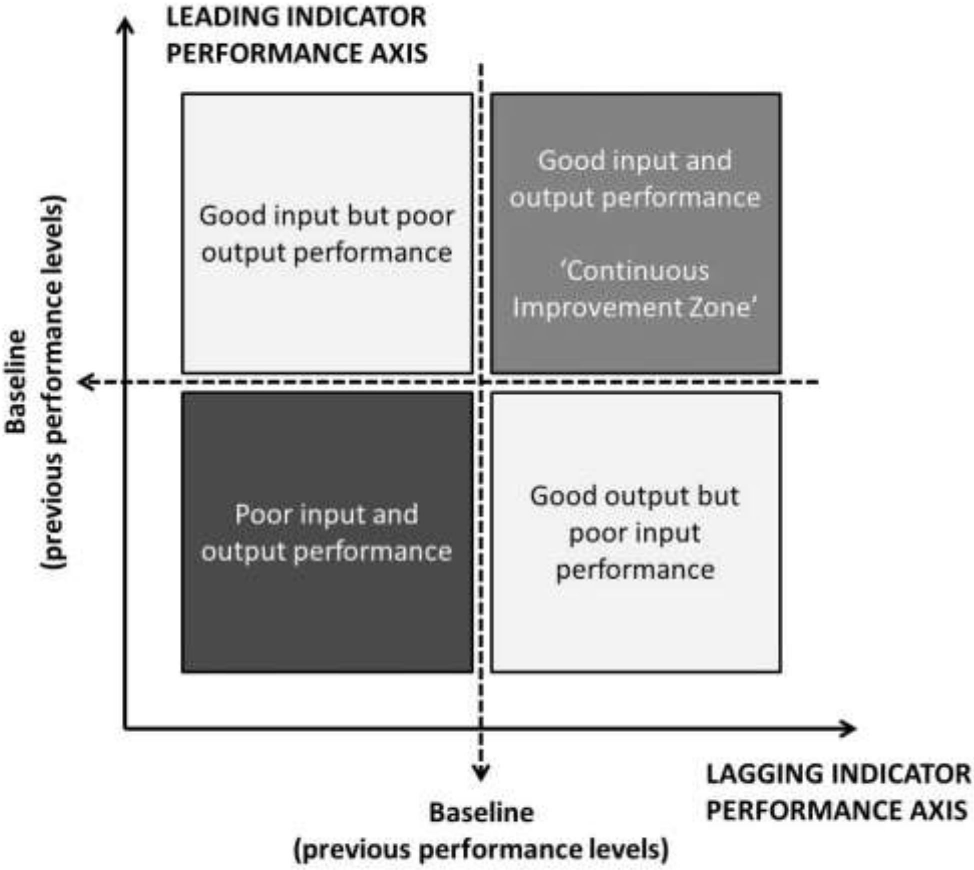

Figure 4 – A matrix is used in the oil and gas industry to present a 'system overview' of safety performance indicators. Defined leading and lagging indicators (such as incident rates, training activities, management visits etc.) are used as coordinates to represent safety performance visually (Source: Step Change, 2003).

An approximate analysis of the UK aviation sector would use fatal accident rate as a lagging indicator and non-fatal incidents as a leading indicator. In the aviation sector an 'accident' is formally defined as "a person suffers fatal or serious injury as a result of [..] being in or upon the aircraft" and/or "the aircraft sustains damage or structural failure" (CAA, 2011, p. A1.1). Incidents and occurrences are in descending order of seriousness. Whilst in the period 2000/2009 there were zero fatalities associated with UK passenger flights (over 5.7 tonnes) there were 179 incidents and 49,000 occurrences. , Unlike accidents, occurrences show a 20% increase over the same ten year period (CAA, 2011). Given that the horizontal/vertical dividing lines in the matrix (Figure 4) are defined by historical baselines, the UK aviation sector would seem to fall on or slightly below the dividing line between Quadrant 4 and Quadrant 2. The point to note is that *"Quadrant 2 represents poor input performance* [relative to a historical baseline] *but good output performance. As there is always an element of luck as to whether inadequately controlled hazards actually result in harm,* *performance in this quadrant represents an organisation that has been lucky. However, one cannot rely on luck. Future performance may not be so good and more attention to input is required to maintain the performance"* (Step Change, 2003, p. 10). If the aviation sector, a leader in safety science, are skirting close to the 'more attention to input required' quadrant, then other sectors are likely to require even closer attention.

Risk Triangle

Underlying the notion of learning from incidents via leading and/or lagging indicators is an even more fundamental safety science concept that is worth revisiting: the risk triangle. In its original form (e.g. Heinrich, 1941) it proposes that 300 recordable accidents will give rise to 30 major accidents, a ratio of 10/1. These proportions were based on data from real accidents in the 1930's. Clearly, in the intervening eighty years the context in which accidents occur has changed dramatically and so too have the proportions expressed in the risk triangle. Indeed, the original Heinrich risk triangle was designed to express the risks to the same person performing the same task in the same environment, and is thus based around individual exposure. The triangle was updated in the 1960's to better express population level risks. Once again, the underpinning analysis was performed within the insurance industry (this time by Frank Bird; 1969) and was based on 1.75 million accidents reported by 297 participating companies across a wide range of sectors. Instead of the 1-30-300 ratio proposed by Heinrich, the revised triangle saw the base broadening to 600 and a 20/1 ratio between near miss incidents and injury accidents. The 1969 risk triangle is again caveated by the methodology used to construct it. In this case the incidents that occurred at the bottom of the risk pyramid were derived from confidential interviews meaning there is a potential data quality and resolution issue. As per the Heinrich risk triangle there will be inevitable domain differences, and again, it represents the safety situation pertaining in the 1960's/1970's. These differences are revealed within the petrochemical industry, early adopters of safety management principles, where more recent research (i.e. Prem & Mannan, 2010) shows there are over 30,000 no injury accidents compared to 961 major injuries: a ratio of 170/1 (compared to Heinrich's original ratio of 10/1). In the UK aviation industry the base of the risk triangle appears even broader, with 49,000 'occurrences' yet only 179 'serious incidents' (a ratio of 274/1).

As risk pyramids become broader they amplify a point made by Heinrich in 1941, that a focus on major accidents (i.e. lagging indicators) "was misdirected and caused valuable data [leading indicators at the base of the pyramid] to be ignored" (Coury et al., 2010). Bird

(1969) goes on to state explicitly that the purpose of these triangles is not to propose a universal set of ratios between levels but to highlight the broader principle that for every major incident there are many more pre-cursor events. The key point relevant to this Special Issue is the intersection between risk pyramids with very wide ‘bases’ of non-injury accidents, and the technological feasibility of collecting large quantities of data via new recording and logging technologies. An intersection exploited by a process not routinely reported in this journal yet one highly relevant to Learning from Incidents. The process is called Flight Data Monitoring.

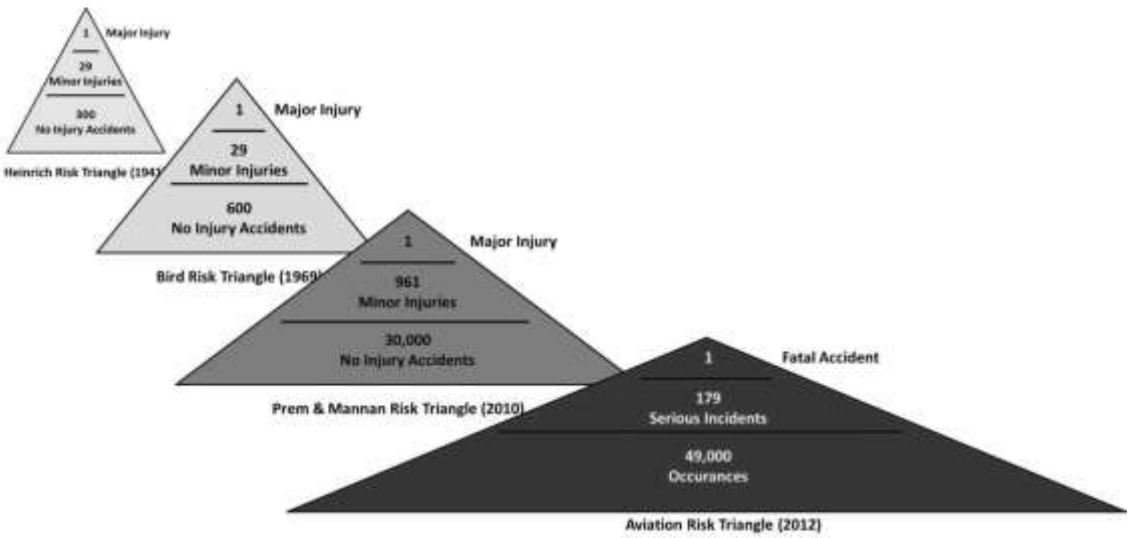

Figure 5 – The original 1941 risk pyramid (Heinrich), the 1969 revised version (Bird), and contemporary risk triangles derived from the petro-chemical (Prem & Mannan, 2010) and aviation industries. Whilst the shape and data ratios vary the principle of a broad base of precursor (leading indicators) underneath a much smaller number of serious or fatal accidents (lagging indicators) remains.

FLIGHT DATA MONITORING

Definition

The impetus for the development of data and voice recording in the aviation domain was the wish to learn from incidents with the data used as a lagging indicator. Having delivered significant safety improvements the relative infrequency of these events now limits the insights a traditional ‘learning from incidents’ approach can offer, (e.g. Baxter, 1995; Hovden, Storseth & Tinmannsvik, 2011; Lundberg, Rollenhagen & Hollnagel, 2009; Pomeroy &

Earthy, 2017etc). To address these limitations there has been a move towards identifying incidents that have the potential to lead to accidents and incidents. A safety science approach highly consistent with the fundamentals of leading indicators and risk triangles (e.g. Heinrich, 1941; Bird, 1969) this is likely to be instructive for numerous other domains too (e.g. Pomeroy & Earthy, 2017). Flight Data Monitoring (FDM) is the process of extracting and analysing data from on-board recorders after *every* journey, with the aim of detecting subtle trends that may help identify potential issues. FDM is defined as:

"A systematic method of accessing, analysing, and acting upon information obtained from digital flight data records of routine flight operations to improve safety. It is the pro-active and timely use of flight data to identify and address operational risks before they can lead to incidents and accidents." (CAA, 2003, pp 1-2).

Recent developments in FDM have drawn on insights offered by analysis of this routine data to inform aviation policy and procedure in areas such as fuel economy, maintenance, passenger comfort, certification of new equipment and training, amongst others. According to one major airline: "The flight safety analysis benefits are proven. The technical benefits have been shown to pay back the original investment in FDM many times over" (Pouliot, 2012). The accumulation of marginal benefits can be considerable when considering a fleet of aircraft operating intensively year-round. FDM driven changes to operational procedures at one airline have reduced the total fuel bill by 1.5% (Pouliot). Another example is fuel burn during 'gate waits', i.e. the aircraft taxis to its designated stand but has to wait for 'operational reasons'. 'Gate waits' in a twelve month period at one airport amounted to 32 hours and over $22,000 of fuel (Pouliot). FDM analysis has enabled the airline to 'monitor system efforts to reduce these costs'. Solutions based on such monitoring include single engine taxiing (reducing fuel usage by approx. 30% over a 45 minute taxi run) and analysis of the costs and usage of auxiliary power units (Pouliot). As already noted, cost savings of this sort often render the safety benefits of FDM effectively free of charge.

Structure and Objectives of FDM System

Under ICAO Annex 6 Part 1 (Amendment) flight data analysis is mandatory for operators of aeroplanes of a certified take-off mass in excess of 27,000 kg. In the UK guidance on the implementation of this directive is provided by the Civil Aviation Authority (CAA) in document CAP 739. Other nations have similar guidance but the development of FDM has been pioneered by the CAA since the 1970's in conjunction with a major UK airline. The CAA suggest that FDM should form part of a feedback loop, preferably as part of a Safety Management System (SMS), and should be constructed with the following objectives in mind:

1. To identify areas of operational risk and quantify current safety margins. e.g. current rates of rejected take-offs, hard landings, unstable approaches, etc.
2. Identify and quantify changing operational risks by highlighting when non-standard, unusual or unsafe circumstances occur. e.g. increases in above rates, new events, new locations.
3. To use the information on the frequency of occurrence(s), combined with an estimation of the level of severity, to assess the risks and to determine which may become unacceptable if the discovered trend continues. e.g. a new procedure has introduced high rates of descent that are approaching the threshold for triggering ground proximity warnings.
4. To put in place appropriate risk mitigation techniques to provide remedial action once an unacceptable risk, either actually present or predicted by trending, has been identified. Once an unacceptable risk has been identified appropriate risk mitigation techniques should be employed to remedy this.
5. Following remedial actions, the situation should continue to be monitored to ensure that the action was effective in targeting the identified risk and that it did not transfer the risk to another area of operation.

Principles of the FDM system

Defining Normality

To enable operators to detect when activities deviate from acceptable practice it is first necessary to define 'what is normal?' Routine collection of data from aircraft in standard operation allows operators to measure normality and establish a baseline. It is then possible to explore the causal relations between input and output variables in order to define explicit leading indicators. For example, approach speed, stability and overall 'energy' is a valuable leading indicator for runway excursions. Pitch rates and flap settings serve as leading indicators for loss of control events and so on.

Event Detection

The traditional approach to FDM is focused on exceedance or event detection. According to CAP 739 events are defined as:

“deviations from flight manual limits, standard operating procedures and good airmanship” (CAP 739, p16).

Computer software can be employed to automatically scan FDR data for instances of these deviations and a set of core events that cover the main areas of interest are quite standard across operators. These events typically include the following:

- High take-off rotation rate
- Stall warning
- Ground proximity warning
- Flap limit speed exceedance
- Fast approach
- High/low on glideslope
- Heavy landing

A more detailed breakdown of the typical exceedance event set is provided in CAP 739, Appendix B. Event detection itself is commonly based on fairly simple statistical techniques and automatic algorithms that detect different phases of flight. In practice, the distribution of events, in most cases, appears to be approximately normal in nature. The standard deviation (SD) over the previous six months is calculated and a ‘traffic light’ system is often used to classify the findings. Results within expected levels are green, those with one or more SD outside the expected range are given an amber (Alert) classification and are flagged for further investigation. Those with three or SDs from expected results are given a red (Action) classification, requiring immediate investigation and remedial action (Henchie, 2011). To give a feel for the rate of event detection, bearing in mind these events will, for the most part, be unnoticeable from a passenger point of view, an airline with a fleet of over 200 aircraft will detect in the region of 7000 red events per month using this method.

Considerable opportunity exists to extend the statistical rigour and sophistication of event detection methodologies and is an area in which further safety science work could derive benefits.

Validation

While the basic statistical approach used in FDM allows operators to understand what constitutes normal performance, insights from manufacturers, industry norms and expert opinions also help define what is acceptable. When the FDM process has correctly identified that aircraft are operating outside normal or acceptable limits, the findings must be interpreted in order to establish which of these represent uncontrollable fluctuations due to external events (e.g. due to weather conditions) and which require remedial action. This step is referred to as validation and recognizes the highly contextual and sometimes nonsensical nature of the data samples obtained. A good example is provided by one airline who, during the month of January, identified high levels of both runway excursions and ‘speed high at touchdown’ events. These events were interpreted as being related and the increases were linked to unusually high winds in the UK during that particular month. The flag raised by these events identifies the issues and further investigation allows actions to be taken to address them, if required. For example, by examining historical records it is possible to expect an increase in these events and implement controls to limit these seasonal effects (Henchie, 2011).

Monitoring

After events have been identified and remedial actions employed to address them, monitoring must continue to ensure the actions were effective in targeting the identified risk and also that actions did not merely transfer risk to another area of operation. For example, if changes in approach procedures are introduced to an airfield with the aim of reducing landings with high rates of descent monitoring other measures at the airfield ensures these do not introduce new risks, such as runway overruns. This is a highly desirable ‘systemic’ aspect of the FDM approach. For example, an airline identified high levels of ‘long landings’, which are one of the factors linked to runway excursions. The airline undertook education of the flight crew to reduce these incidents and measured the effect of their intervention. The evidence supported the effectiveness of the training in terms of the average touchdown position with a reduction in long landings. However, altering the standard operating procedures to emphasise the importance of touching down in the designated touchdown area had an unanticipated consequence. There was a significant increase in ‘low level go arounds’. The intervention had indeed shifted the problem to another area of operation which could be investigated further (Henchie, 2011).

Confidentiality and Safety Culture

FDM is part of a broader safety culture in which the flight crew are encouraged to report operational issues, events and potential problems. The aviation sector is far from immune to industrial relations challenges and has spent several decades managing them in relation to FDM and data recording more generally. The aviation sector provides a useful lesson in how these challenges can be surmounted. In successful FDM implementations ‘crew feel comfortable reporting safety incidents with the knowledge that the information will be used to improve safety rather than punish transgressions.’ (Henchie, 2011). Indeed, it is not unheard of for crew to remark that something “will trigger the FDM” and, more importantly, that something positive will follow.

The FDM Cycle in Practice

The context of safety investigation practices is important (Rollenhagen et al., 2010). Figure 6 illustrates the current state of the art in FDM, and the processes used to learn from incidents, in a highly systemic manner using leading indicators. In practical terms, the flight data is downloaded from the aircraft and is subject to analysis by an FDM operative who will apply exceedance detection software to the data and will verify the events detected. An FDM analyst will assess the findings and will feedback to the crew (via the Trade Union) and to safety officers at monthly review meetings. Insights gained from FDM analyses will be combined with those from air safety and other reports to form a clearer picture of safety issues. The outcomes of thes meetings will feed into a number of other channels; trend information will be reported to management and staff; any new safety issues identified will be used to inform flight data investigations and will contribute to improvements to the monitoring software; insights will also be applied to practice via changes to operating procedures, manuals and crew training. These changes will impact on the FDM process itself serving to feed a process of continuous improvement.

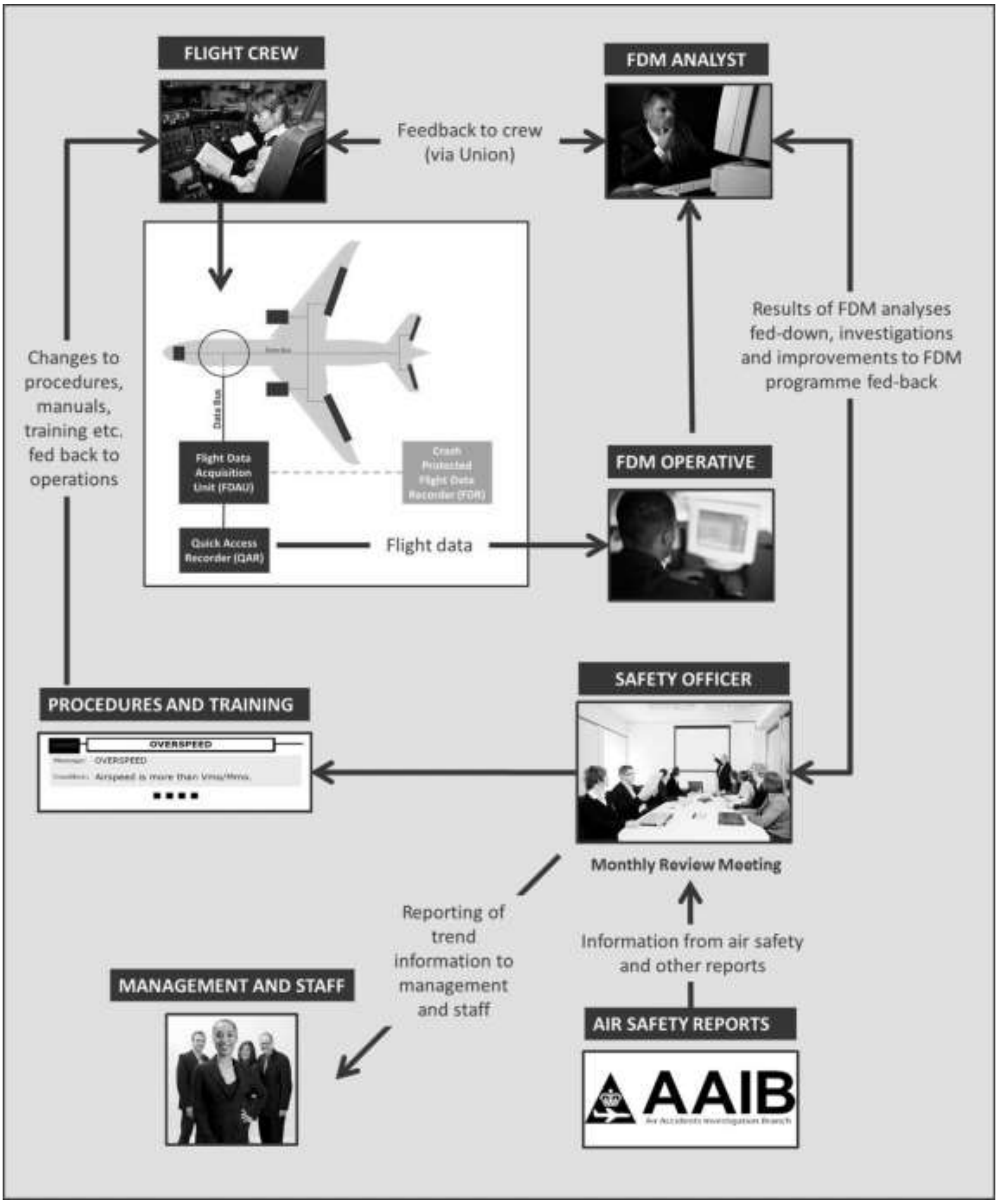

Figure 6 – Illustration of the 'FDM cycle'

CONCLUSIONS

2010 represented a landmark year in European aviation safety. This was not just because it was the first year in the entire history of European aviation that no fatal commercial air transport incidents occurred at all, but because it signals a paradigm shift in how incidents should be learnt from. The aviation sector has for most of its life been a leader in safety science and continues to be so. By reviewing the safety impact of ‘black box recorders’ a significant paradox is revealed: despite an abundance of data there are now vanishingly few opportunities to use it to ‘learn from incidents’ simply because there are so few to learn from. This inverse relationship between quantity of data and quantity of major incidents gives rise to three key safety science issues that should shape the future of the discipline in theory and in practice.

The first key issue is that never before in history has it been possible to acquire so much information on so many operational parameters. This is as true for the aviation sector as it is for countless others. In the vast majority of cases these stores of data represent an underused source of operational and safety insight. The lesson from aviation is that a comparable ‘Data Monitoring programme’ is applicable to many other domains which now routinely collect operational data Even within the aviation sector there are considerable opportunities to enhance the statistical and analytical rigour of event detection and validation. Moreover, there appears to be a major opportunity to couple processes such as FDM to the emerging field of Big Data Analytics, particularly as it relates to human performance issues (Walker & Strathie, 2016; Walker et al., In Press; Drury, 2015). Going further still, it is very apparent that many strategic safety risks are couched firmly in Human Factors (see CAA, 2011) and that the outputs of data monitoring could feed directly into Human Factors Methods (see Stanton et al., 2013) in order to drive out the much sought after behavioural insights. This work is already underway by the authors (see Walker & Strathie, 2015; 2016; Strathie & Walker, In Press).

The second key issue is that it has been known for a long time that lagging indicators represent a relatively poor source of safety insight, both practically and ethically. By returning to this fundamental literature, and reviewing the concepts of leading and lagging indicators in depth, it is possible to issue a reminder that a better way exists. The data which is routinely and increasingly

collected from safety critical systems could be used to learn from ‘incidents’ rather than from ‘disasters’. The confluence of risk pyramids that now have extremely broad bases and narrow tips and the amount of data now available on normal operations, suggests a re-definition of what constitutes an ‘incident’. Previously, it was only lagging indicators that could provide the forensic level of detail required to define this. With data recording in depth data can be gathered from normal operations in a manner not previously possible. By leaning on the concepts of leading indicators it is clear that 'incidents' could be things that would otherwise be unnoticeable: not just 'weak signals' (e.g. McLeod, 2015) but perhaps 'nearly imperceptible signals'.

The third key issue is a firmly practical one. Many of the antecedent conditions for comparable Data Monitoring process are in place within numerous sectors: in many cases the data acquisition architecture is present (e.g. SCADA systems), the modern, solid-state data collection hardware is now mandatory (e.g. On Train Monitoring and Recording) and many of the safety drivers to be found in aviation are now just as critical elsewhere (e.g. a plateauing of safety performance and an increasing focus on behavioural issues and Human Factors). It is hoped that by presenting this example of state-of-science, relating it to fundamental principles in the discipline and revealing the considerable opportunities will help to bring forward a new research agenda . At the very least, this review prompts us to question – at a fundamental level – what learning from incidents really means and what it could look like in practice.

ACKNOWLEDGEMENTS

This work was funded by the UK Economic and Physical Sciences Research Council (EPSRC) under Grant Reference: EPSRC EP/I036222/1. The involvement and support of the UK Civil Aviation Authority, in particular David Wright and Joji Waites, is gratefully acknowledged.

REFERENCES

AAIB (2011). Annual Air Safety Report. Retrieved 24 July 2012 from: www.aaib.gov.uk.

AIB (1973). Trident I G-ARPI: Report of the Public Inquiry into the Causes and Circumstances of the Accident near Staines on 18 June 1972. London, HMSO.

ARINC (2016a). 573-7 Aircraft Integrated Data System Mark 2 (AIDS Mark 2). Annapolis, MD; ARINC.

ARINC (2016b). 717-15 Flight Data Acquisition and Recording System. Annapolis, MD; ARINC.

ARINC (2016c). 767-1 Enhanced Airborne Flight Recorder. Annapolis, MD; ARINC.

Baxter, T. (1995). Independent investigation of transportation accidents. Safety Science, 19, 271-278.

Bird, F.E. & Germain, G.L. (1985) Practical Loss Control Leadership. Loganville, GA: International Loss Control Institute Incorporated.

CAA (2003). CAP 739: Flight Data Monitoring: A Guide to Good Practice. London, CAA.

CAA (2008). CAA Safety Regulation Group: Safety Plan Update 2008. UK: Civil Aviation Authority.

CAA (2011). Safety Plan 2011 – 2013. London, Civil Aviation Authority.

Campbell, N. (2007). The Evolution of Flight Data Analysis. In: ASASI 2007: Proceedings of Australian Society of Air Safety Investigators Conference, Sydney, Australia. Retrieved 10th Jan 2016 from: http://asasi.org/papers/2007/The_Evolution_of_Flight_Data_Analysis_Neil_Campbell.pdf.

CIAIAC (1999). Technical Report A-054/1999: Accident to Boeing 757-200 G-BYAG at Girona Airport on 14 September 1999. Madrid, Spain: CIAIAC.

Coury, B. G., Ellingstad, V. S. & Kolly, J. M. (2010). Transport accident investigation. The development of human factors research and practice. Reviews of Human Factors and Ergonomics, 6(1), 1 – 33.

Defence Science and Technology Organisation (Australia) (2005). The History of the Black Box. Retrieved 12 July 2012 from: http://www.dsto.defence.gov.au/attachments/The%20Black%20Box.pdf

Drury, C. (2015). Human factors/ergonomics implications of big data analytics: Chartered Institute of Ergonomics and Human Factors annual lecture. Ergonomics, 58(5), 659-673.

European Aviation Safety Authority (2010). Annual Safety Review 2010. Cologne, Germany: EASA.

Ford, M (2012). Use of Data Mining Techniques During the Investigation of the Investigation of Boeing 777-236ER, G-YMMM London Heathrow Airport, 17 January 2008. Presented at FDM Short Course, Cranfield University, UK.

Heinrich, H.W. (1941) Industrial Accident Prevention: A Scientific Approach (2nd Ed.). New York: McGraw-Hill.

Henchie, S. (2011). Flight Data Monitoring within a Safety Management Programme. Presented at FDM Short Course, Cranfield University, UK.

Hinze, J., Thurman, S. & Wehle, A. (2013). Leading indicators of construction safety performance. Safety Science, 51(1), 23-28.

Hovden, J., Storseth, F. & Tinmannsvik, R. K. (2011). Multilevel learning from accidents: case studies in transport. Safety Science, 49(1), 98-105.

Hutton, G. B. (1972). Civil Aircraft Airworthiness Data Recording Programme: A Study of Normal Operational Landing Performance on Subsonic Civil Jet Aircraft. London, HMSO/Aeronautical Research Council.

King, G. E. (1967). Civil Aircraft Airworthiness Data Recording Programme: Special Events of Meteorological Origin (January 1966 to November 1968). London, HMSO/Aeronautical Research Council.

Lundberg, J., Rollenhagen, C. & Hollnagel, E. (2009). What-you-look-for-is-what-you-find – The consequences of underlying accident models in eight accident investigation manuals. Safety Science, 47(10), 1297-1311.

McLeod, R. W. (2015). Designing for Human Reliability. Houston: Gulf Professional Publishing.

Morcom, A.R. (1970). Flight data recording Systems: A brief survey of the past developments, current status and future trends in flight recording for accident investigation and operational purposes. Aircraft Engineering and Aerospace Technology, 42 (7)12 – 16.

Pomeroy, R. V. & Earthy J. V. (2017). Merchant shipping's reliance on learning from incidents – A habit that needs to change for a challenging future. Safety Science, Available online 13 February 2017

Pouliot, R. (2012). Expanding the uses of FDM - Airlines. Presented at FDM Short Course, Cranfield University, UK.

Prem, K.P., Ng, D., Mannan, S.S., (2010). Harnessing database resources for understanding the profile of chemical process industry incidents. Journal of Loss Prevention in the Process Industries 23, 549 - 560.

Rail Safety & Standards Board (RSSB) (2004). Review of the efficacy of assurance process in preventing catastrophic accidents, in management. London: Rail Safety and Standards Board.

Rogers, A., Evans, R., and Wright, M. (2009). Leading Indicators for Assessing Reduction in Risk of Long Latency Diseases. Greenstreet Berman Limited for Health and Safety Executive.

Rollenhagen, C., Westerlund, J., Lundberg, J. & Hollnagel, E. (2010). The context and habits of accident investigation practices: A study of 108 Swedish investigators. Safety Science, 48(7), 859-867.

Schminke, L. H., Jeger, V., Evangelopoulos, D. S., Zimmerman, H., Exadaktylos, A. K. (2013). Riding the escalator: How dangerous is it really? Western Journal of Emergency Medicine, 14(2), 141–145.

Shea, T., De Cieri, H., Donohue, R., Cooper, B. & Sheehan, C. (2016). Leading indicators of occupational health and safety: An employee and workplace level validation study. Safety Science, 85, 293-304.

Sinelnikov, S., Inouye, J. & Kerper, S. (2015). Using leading indicators to measure occupational health and safety performance. Safety Science, 72, 240-248.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C. & Jenkins, D. P. (2013). Human factors methods: a practical guide for engineering and design (2nd Edition). Farnham: Ashgate.

Step Change in Safety (2003). Leading Performance Indicators Guidance for Effective Use Brochure. Aberdeen, UK: Step Change in Safety.

Strathie, A. & Walker, G. H. (In Press). Applying link analysis to train recorder data in order to identify patterns of behaviour. Human Factors.

Walker, G. H., Salmon, P. M., Bedinger, M. & Stanton, N. A. (Under Review). Quantum ergonomics. Ergonomics: Special Issue on New Paradigms.

Walker, G. & Strathie, A. (2015). Leading indicators of operational risk on the railway: A novel use for underutilised data recordings. Safety Science, 74, 93-101.

Warren, D. [interviewee] (2004). A Major Contribution to the Enhancement of Air Safety. [Interview transcript]. Retrieved 18 July 2012 from: http://www.iasa.com.au/folders/Safety_Issues/dfdr-cvr/drWarren.html

Warren, D.R., (1954). A Device for Assisting Investigation into Aircraft Accidents. ARL-MECH-ENG-TECH-MEMO-142.

World Health Organisation (2016). Regional Office for Europe: European Mortality Database. Retrieved 29th Feb from: http://data.euro.who.int/hfamdb.

Wright, D. (2012a). FDM Regulatory Framework and Advisory Material. Presented at FDM Short Course, Cranfield University, UK.

Wright, D. (2012b). Principles of Data Validation & Assessment. Presented at FDM Short Course, Cranfield University, UK.