

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

THE DEVELOPMENT OF A MORE RISK-SENSITIVE & FLEXIBLE AIRPORT SAFETY AREA STRATEGY: Part I The development of an improved accident frequency model

D.K.Y. Wong, D.E. Pitfield¹, R.E. Caves, A.J. Appleyard
Transport Studies Group, Civil & Building Engineering Dept.
Loughborough University, Loughborough, LE11 3TU U.K.

Abstract

This two-part paper presents the development of an improved airport risk assessment methodology aimed at assessing risks related to aircraft accidents at and in the vicinity of airports and managing Airport Safety Areas (ASAs) as a risk mitigation measure. The improved methodology is more quantitative, risk-sensitive, flexible and transparent than standard risk assessment approaches. As such, it contributes to the implementation of Safety Management Systems at airports, as stipulated by the International Civil Aviation Organisation.

The first part of the paper presents the methodological advances made in the development of accident frequency models; namely the building of a single comprehensive database of all relevant accident types, the collection and use of normal operations data in quantifying the criticality of a series of risk factors, and modelling accident frequency using multivariate logistic regression. The resulting models have better goodness-of-fit, sensitivity and specificity than standard risk assessment methodologies.

1. Introduction

International as well as local aviation authorities have developed airport safety areas (ASA) at and around airports to protect passengers as well as nearby communities from accidents that occur during the take-off and landing phases of flight. ASAs could be grouped into two families – aerodrome design ASAs and land-use planning ASAs.

¹ Corresponding author, E-mail: D.E.Pitfield@lboro.ac.uk

In terms of aerodrome design ASAs, there is an internationally agreed framework on airport design set out in Annex 14 to the Convention on International Civil Aviation (ICAO 1999). National aviation authorities, however, may deviate from Annex 14 or develop different standards. For instance, the FAA's Advisory Circular 150/5300-13 on Airport Design is a parallel framework to Annex 14 (FAA 2004), as is the UK's CAP 168. The notion of ASAs, therefore, tends to differ from country to country. Under ICAO Annex 14, safety areas relevant to take-off and landing accidents include the Runway End Safety Area and Runway Strip. The concepts of the Runway Strip and the RESA are combined under FAA rules, which define the Runway Safety Area. The FAA also specifies a Runway Protection Zone, which has no equivalent in ICAO Annex 14.

Land-use planning ASAs result from regulations and guidelines that govern the way land is used around runways. There are relatively few national regulations on land-use near airports, let alone an international framework. The most notable jurisdictions that have instituted land-use planning ASA include the Netherlands, the United Kingdom and certain states in the US such as California. The regulations concerned are often formed from the concept of risk contours and prohibiting development within.

Wong (2007) highlighted a number of fundamental deficiencies concerning Airport Safety Area (ASA) regulations. These include the number of risk factors considered in the formulation of ASA policies; their rigid, prescriptive and compartmentalised nature; opacity in rule-making and the lack of review mechanisms; a piecemeal and reactive approach; a fragmented oversight regime; a "tick the box" compliance mentality on behalf of the regulated parties; and the overall regulatory rationale. Above all, current ASA requirements stipulate average levels of safety across vastly different airports, contributing to a significant mismatch between actual risk exposure and safety margin provision. The need for a more risk-sensitive, flexible and effective strategy of using and regulating ASAs is clear. As the first of a two-part paper, this paper presents the development of an accident frequency model that would be central to the improved utilisation and requirements of ASAs. The frequency model in an ASA-related risk assessment considers the probability of an accident

occurring in the vicinity of an airport. This follows established practice of risk assessment in the field (Piers 1996, DfT 1997, Hale 2002).

The following section puts forward the methodological advances made in this paper, namely an integrated approach, the building of a single comprehensive accident database, the expanded use of Normal Operations Data (NOD) and the inclusion of new risk factors. The first two advances guided the way the accident database was developed (described in Section 3) and the use of new sources of NOD (described in Section 4) allowed the inclusion of the additional risk factors.

2. Advances in Methodology

The model developed offers a new approach to accident frequency modelling addressing some key deficiencies of current risk mitigation measures and risk assessment methodologies as described in Part I of this paper. These advances were made possible by expanding the traditional scope of airport risk assessment studies, building comprehensive and compatible accident and normal operations databases and developing multi-dimensional quantitative models that explicitly take into account previously neglected risk factors. These are detailed below.

2.1 Integrated approach

This research takes an integrated approach to airport risk assessment rather than focusing on a single stakeholder or element of the aviation system. The study crosses existing regulatory boundaries and considers aircraft crash risk on both sides of the airport fence, reflecting the geographically continuous nature of accident risk. This facilitates complementary policies in aerodrome design, land-use planning and operational parameters to be developed in lieu of the current fragmented and compartmentalised risk control measures. It has never been done before and avoids the difficulties of drawing from studies with different objectives and assumptions. The need for such an approach is evidenced in the responses to the New Zealand Civil Aviation Authority's consultation on its Runway End Safety Area (RESA) policy where respondents suggested that more aerodrome physical requirements be assessed along with the RESA in a single coherent study (Watson 2005).

2.2 Single comprehensive database

Another advance made by this study is the comprehensive accident database developed. Unlike previous studies that focused on a specific type of accident, such as approach-and-landing accidents (Enders et al. 1996, Khatwa & Helmreich 1998), third-party accidents (DfT 1997) or overruns (CAA 1998), all accident types that are implicated by ASAs are included in this study – take-off and landing overruns, undershoots, veer-offs as well as crashes after take-off. This facilitates the assessment of all accident types in a coherent manner, rather than being based on multiple databases with different inclusion criteria. All accident types are sampled from the same period and for the same parameters using a set of standardised rules. More definitive conclusions on ASA policies could therefore be drawn. For example, Kirkland’s work (Kirkland et al. 2003) considered overruns but not undershoots or crashes after take-off. Having included the latter two types of accidents for modelling, the current study provides the complete analysis of RESA and (Public Safety Zone) PSZ needs.

2.3 Normal operation risk exposure

Another methodological advance is the use of normal operations (i.e. non-accident flight) data for risk modelling, specifically data related to flight operations and meteorological conditions. Various studies have already identified the lack of normal operations data (NOD) as a major obstacle to the development of quantitative risk models (DOT 1979, Piers et al. 1993, Khatwa et al. 1996, Khatwa & Helmreich 1998, Eddowes et al. 2001, Li et al. 2001). For example, a NLR study on the impact of crosswind on aircraft operations noted that “the significance of [risk] factors can only be established when the number of non-accident flights, under identical circumstances is known” (Van Es et al. 2001). Enders et al. (1996) stated that the unavailability of NOD hampered the calculation of accident occurrence rates and the ICAO concurs that the absence of NOD “compromises the utility of safety analysis” (ICAO 2006). Indeed, in the absence of information on risk exposure, even though the occurrence of a factor, e.g. contaminated runway, could be identified as a contributor to many accidents, it is impossible to know how critical the factor is since many other flights may have also experienced the factor without incident. With NOD, the number of operations that experience the factor singly and in combination with other factors could be calculated, so risk ratios could be generated and the importance of risk

factors quantified. This would allow the allocation of resources for safety improvement to be prioritised (Enders et al. 1996).

This paper represents a step forward in the field of airport risk assessment in collecting a large and representative sample of disaggregate NOD covering a range of operational and meteorological risk factors, allowing their criticality to be quantified. Incorporating this risk exposure information into the accident frequency model enhances its predictive power and provides the basis for formulating more risk-sensitive and responsive ASA policies. Accident frequency models need no longer rely on simple crash rates based on just aircraft, engine or operation type. As discussed below, factors previously ignored by airport risk assessments and ASA regulations are accounted for using the models developed in this study. Moreover, this normal operations database is not only valuable for the current project but can also be used for future studies.

2.4 Factors considered

In addition to airline Flight Operational Quality Assurance (FOQA) or Flight Data Recorder (FDR) data through which airlines use to monitor aircraft performance, only in human factor and crew resource management analysis is the use of NOD relatively established. Khatwa and Helmreich (1998) used Line Operations Safety Audits (LOSA) to analyse crew errors during non-accident flights. Work at the University of Texas at Austin (Helmreich et al. 1999, Klinec et al. 1999) also used LOSAs to build conceptual models that represent the operating environment. Beyond human error analysis, the use of NOD in risk assessment is limited, especially for airport-related risks. Enders et al. (1996) and Roelen et al. (2000) used aggregate NOD to establish risk ratios for various risk factors such as the availability of Terminal Area Radar and other airport navigational aids. Many attempts to incorporate NOD in risk assessment failed because the available risk exposure data does not allow subdivision in movements based on the risk factors of interest (Piers 1994, 1998). Kirkland et al (2003) broke new ground in the use of disaggregate NOD for assessing aircraft overrun risk. Using a limited sample of NOD, three overrun risk models were built. Two of them assessed overrun risk based on aircraft weight as a percentage of the maximum take-off and landing weight respectively and the third model considered landing overrun risk based on the distance of excess runway available. Although

some insightful conclusions were drawn, the number of risk factors that could be modelled remained small.

One notable gap in research is the quantification and modelling of the criticality of meteorological risk factors to accident occurrence. The lack of data on flights' exposure to meteorological conditions meant traditional risk assessment had to rely on qualitative judgements (Eddowes et al. 2001) or simply ignore meteorological conditions as risk factors, as do most ASA policies. Although Enders et al. (1996) acknowledged that adverse weather conditions is one of the most regularly cited factors in accident reports, they were unable to include the terms in their analysis. Kirkland also cited the lack of meteorological NOD as a major shortcoming of his work (Kirkland 2001). The current study was able to collect exposure data on a range of meteorological parameters and include them in accident frequency modelling – ceiling height, visibility, crosswind, temperature, fog, precipitation, electric storm, snow, frozen precipitation and icing conditions. Other factors not commonly modelled were also taken into account, e.g. airport hub size, terrain surrounding the airport, dawn and dusk conditions as well as foreign or domestic operation. This is in addition to the more traditional parameters of aircraft, engine and operation type. The current paper is thus able to provide a far more comprehensive analysis of risk factors relevant to airport risk assessment and develop state-of-the-art frequency models of accident occurrence covering an unprecedented spectrum of risk factors.

3. Accident Database

The low accident rate of aviation means that no particular airport has sufficient accident occurrences in the recent past to support an accident frequency model with reasonable statistical confidence (Piers et al. 1993, Piers 1994, Hale 2001). Therefore, a robust risk assessment must draw from a large database of relevant accident cases.

Accident types commonly associated with the safety areas include overruns, veer-offs, undershoots, crashes after take-off and third party accidents. However, these accident classifications are based more on consequence than cause. For example, third party accidents may simply be undershoots or crashes after take-off if no third parties were present. Similarly, an aircraft overrunning the runway end and another

veering off the side of the runway are often just different manifestations of the same root problem, differing only in crash kinematics and airfield conditions. The database classes take-off and landing related accidents under four categories – landing overruns (LDOR), landing undershoots (LDUS), take-off overruns (TOOR) and crashes after take-off (TOC). This classification essentially separates landing and take-off accidents then ground-based and airborne accidents. This facilitates their analysis by cause rather than consequence, which is especially appropriate for developing the accident frequency model. Incorporating third party accidents within these four accident types instead of considering them in isolation reflects the geographically continuous nature of accident risk. Keeping the number of accident categories to a small number also helps to increase the statistical significance of subsequent analyses and models. Only cases in which at least one ASA was directly challenged, or impacting ground or an obstacle within 10km of the landing or take-off threshold were included in the database. Directly challenging the ASA means that the aircraft has exited from the ‘normal’ areas of operation on the airfield, e.g. veering off the runway or hitting obstacles on landing or take-off. Subsequent location analysis confirms that the 10km cut-off has captured the great majority of relevant landing and take-off occurrences.

The feasibility of including relevant incidents not resulting in hull loss was also explored. However, the quantity and quality of data available on these occurrences of lesser consequence are by far inferior to that collected for accidents. Including them would result in a database with a vast number of missing fields. They were therefore not included in the final database.

The US represents the largest national aviation system in the world in terms of air traffic, aircraft and airfields. The US NTSB is also the largest aviation accident investigator with an established database of accidents that have been investigated in a relatively systematic and consistent manner. The preference for studying US aviation safety data has been echoed by other studies such as Button & Drexler (2006). The database therefore consists of all relevant cases that took place in the US between 1982 and 2002.

The data fields of the accident database covered a multitude of parameters including aircraft, flight and airport characteristics, weather conditions, wreckage location and injury levels. The NTSB online accident database alone is not sufficient for the purpose of the current research. Therefore, as with Kirkland (2001), it was necessary to obtain individual accident reports and docket files from the NTSB, even though the amount of information contained in each docket varies greatly. From these sources all available relevant information was extracted. Certain variables required additional calculation based on available data. For instance, crosswind strength was computed from data on wind direction, wind velocity and true runway orientation.

Other than directly challenging at least one ASA or impacting ground or obstacles within 10km of the landing or take-off runway threshold, a number of other criteria were used to filter the NTSB accident database to identify the accidents of interest. These filters, in essence, eliminate cases that involve airports outside the US, irrelevant aircraft types² and operations as well as accidents with minimal consequences. They were developed considering data availability and quality, compatibility with the normal operations data, the need for statistical significance, relevance to large and small airports as well as the criteria used by previous airport risk assessment studies. The final database totalled 440 cases, of which 199 are landing overruns, 122 are landing undershoots, 52 are take-off overruns and 67 are crashes after take-off.

4. Normal Operations Database

The challenges of obtaining appropriate NOD for risk assessment are well documented (Piers 1994, DfT 1997, Roelen et al. 2002). Unavailability, incompleteness and difficult access are only some of the hurdles that must be overcome. A number of sources of NOD were considered for use in the current study. A satisfactory solution was found in the data provided by the FAA's Aviation Policy and Plans Office (APO). The Enhanced Traffic Management System Counts (ETMSC) database provides hourly traffic counts for over 450 airports as well as the relevant traffic characteristics for individual flights, including aircraft, engine and

² These include non-fixed wing aircraft, aircraft with certified maximum gross weight under 6,000lbs, single engine aircraft, piston engine aircraft and FAR Part 91 entries with certified maximum gross weight under 12,500lbs. These aircraft types were removed because they were deemed out of the current research's scope and/or the quality of data available for these entries are poor.

operation type. One of the key advantages of the ETMSC database is that, unlike specific airport or airline FOQA or FDR data, it encompasses a wide variety of airport sizes and includes commercial, air taxi, freight as well as general aviation flights. However, ETMSC does not provide the associated weather and runway orientation information³. Supplementary sources must therefore be used to cover these data gaps.

4.1 NOD Sampling strategy

As it is impractical to sample every non-incident flight during the study period, a sampling strategy must be developed for collecting the appropriate sample of NOD. The prime concern is to gather a sample that is representative of the risk exposure of the overall normal flight population of interest. Extensive effort was spent on sampling appropriately such that the final sample is representative of the non-incident flight population. Random sampling of the ETMSC database would not be appropriate as it may bias against airports of certain risk profiles and misrepresent the genuine risk exposure of normal flights. A stratified sampling strategy was hence developed to select airports from which normal flights were then sampled.

The first stratification factor is airport size (hub and non-hub). This accounts for the difference in risk exposure of flights related to large and small airports including aircraft size, operation type, navigational aid availability, airport infrastructure etc (Piers 1994). The second factor is FAA region, which represents a reasonable division of the key geographical regions of the US. As such, it is a useful stratification factor to account for the broad differences in regional weather patterns and hence normal flights' exposure to various meteorological conditions. The third stratification factor is the presence of significant terrain near the airport as the latter is expected to influence accident risk, especially for landing undershoots. The NOD sample ought to reflect the proportion of flights that are exposed to more challenging topographic environments. An airport is considered to be situated near significant terrain if the terrain within the Instrument Approach Procedure planview exceeds 4,000 feet above the airport elevation, or if the terrain within a 6.0 nautical mile radius of the Airport Reference Point rises to at least 2,000 feet above the airport elevation. Detailed terrain is depicted in the Instrument Approach Procedures of these airports

³ Another APO database, Aviation System Performance Metrics (ASPM), provides meteorological and runway information related to specific flights but its coverage is limited to the relatively large airports.

according to this definition in the FAA US Terminal Procedures Publication (FAA 2007). The two airport classes, nine FAA regions and two terrain categories theoretically lead to 36 strata from which the NOD sample should be drawn.

Table 1 NOD sampling stratification factors

No.	Stratification Factor	Definition
1.	Airport size (Hub/Non-Hub)	Hub airports include all large, medium and small hub airports as classified by the FAA in 2001
2.	FAA Region (9 regions)	See Figure 1
3.	Presence of significant terrain near the airport (Significant Terrain/ Non-Significant Terrain)	Significant if the terrain within the Instrument Approach Procedure planview exceeds 4,000 feet above the airport elevation, or if the terrain within a 6.0 nautical mile radius of the Airport Reference Point rises to at least 2,000 feet above the airport elevation.

Figure 1. FAA regions



However, seven of these never featured among the airports of any FAA Terminal Area Forecasts (TAF) from 2000 to 2005, which comprehensively includes FAA towered airports, federally contracted towered airports, non-federal towered airports as well as non-towered airports. It is therefore reasonable to assume that no airport exists under these strata. An example would be a large hub in the central region in significant terrain. Eliminating these non-existent hypothetical strata resulted in 29 strata with actual airport traffic. The 125 ETMSC airports suitable for NOD sampling were therefore split according to the 29 strata. However, no airport fell under seven of them. The impact of the empty strata was considered before the described sampling strategy was accepted. From TAFs, it was calculated that all airports nationwide belonging to the seven empty strata which cannot be sampled collectively account for 4.2 percent of the total relevant traffic from 2000 to 2005. This figure is considered sufficiently small as to not affect the overall representation of risk exposure of the great majority of relevant normal operations. There remained, then, 22 strata with airports to sample from. If there were five or fewer airports in a particular stratum, all of them were sampled. For strata with more than five candidate airports, five were sampled from each. The five were selected such that airports of different traffic levels are represented. For example, if there were ten airports in the stratum, every other airport would be sampled in the order of descending traffic level. This ensures that the sampled airports correctly reflect the traffic characteristics of the overall normal flight population. This resulted in a total of 78 selected airports. These sampled airports account for 48,924,040 operations from 2000 to 2005 inclusive, i.e. 25.5 percent of all relevant traffic during that period. Clearly, it is impractical and unnecessary to use all operations from these sampled airports for analysis. Therefore only flights on the first day of February, May, August and November of 2002, 2003 and 2004 were sampled to constitute the final NOD sample for risk assessment. The selection of the four months allows seasonal variations in weather exposure to be captured to a degree.

The sampled normal operations data was then filtered as with the accident database to ensure that the two are compatible, i.e. the sampled NOD traffic does not contain

traffic types outside the scope of the accident database. Having eliminated the irrelevant traffic, the final NOD sample consisted of 242,420 flights. Before the sampled NOD could be considered as representative of the overall population of normal operations of interest, the differences in sampling fraction between the 22 strata must be resolved. Sampling every stratum as described above led to certain strata being over-sampled and others under-sampled when compared to the actual composition of the overall normal flight population, since the proportion of available airports for sampling varies stratum to stratum. Proportionate allocation was therefore applied via the use of weights. A specific stratum's weight was derived by dividing the stratum's fraction of traffic in the overall population by the stratum's fraction of traffic in the sampled population. This is similar to inverting the sampling fraction of each stratum but avoids inflating the total number of sampled flights. Inter-strata sampling imbalances were addressed by applying the weights, which calibrated the final NOD sample to reflect the overall normal flight population.

4.2 Supplementary NOD

The ETMSC database provides landing and take-off counts of hourly segments at specific airports broken down by aircraft, engine and operation type. For other risk exposure parameters, additional sources of NOD were found to supplement ETMSC's traffic data.

The National Oceanic and Atmospheric Administration's Integrated Surface Hourly (TD3505) database was selected to measure the sampled ETMSC flights' exposure to a range of meteorological factors. Having collated the appropriate TD3505 data to the relevant ETMSC time segments, it was possible to quantify the normal flights' exposure to a large number of weather parameters. These include visibility, ceiling height, temperature, precipitation, snow, fog, icing condition, electric storm and a host of other weather measures.

Whereas most meteorological conditions were readily identified in TD3505, others required further calculation. TD3505 data on wind direction and velocity was coupled with the true runway orientation of flight operations to compute the

crosswind factor⁴. Another parameter that required further computation was light condition. The accident database recorded whether each incident occurred in daylight, night, dawn or dusk. For normal operations, dawn was defined as the hour before official sunset time and dusk the hour after official sunset. 2002 civil twilight times were used to determine sunrise and sunset hours at locations across the US. According to these designated hours, then, sampled flights that took-off or landed in dawn and dusk were identified. Hours after dusk were identified as night-time and the rest daylight. It is acknowledged that the definitions and methodology used to identify light conditions are somewhat crude and may have overstated the duration of dawn and dusk hours. However, given the hourly time segments of ETMSC, more precise definitions were not possible.

5. Multivariate Modelling

Multivariate statistical models were developed for the prediction of accident occurrence in the context of airport risk assessment. The models take into account the factors captured by both the accident and normal operations databases.

Logistic regression was the preferred statistical procedure for this study for a number of reasons. Firstly the technique is suited to models with a dichotomous outcome (accident and non-accident) with multiple predictor variables that include a mixture of continuous and categorical parameters. Logistic regression is also especially appropriate for case-control studies because it allows the use of samples with different sampling fractions depending on the outcome variable without giving biased results. In this study, it allows the sampling fractions of accident flights and that of normal flights to be different. This property is not shared by most other types of regression analysis (Nagelkerke et al. 2005).

It was ensured that all assumptions for the statistical technique were met. Visibility was entered into the model as a five-level categorical variable in order to meet the logit linearity assumption. Collinearity among the predictor variables was also assessed by conducting linear regression analyses to obtain the relevant tolerance and

⁴ The crosswind factor was calculated by considering the direction of the landing/take-off operation (runway orientation), and the wind direction and velocity at the time of the operation as provided by TD3505. This involved coupling each sampled flight from ETMSC with the relevant reading from TD3505.

variance inflation factor (VIF) values. None of the tolerance values were smaller than one and no VIF value was greater than ten, suggesting that collinearity among the variables is not serious (Myers 1990, Menard 2001). Kendall's tau was also used to assess potential correlations between predictor variables that are likely to be related. Three pairs of variables had Kendall's tau correlation coefficient between 0.5 and 0.65, indicating moderate correlation. They were equipment class with user class, equipment class with airport hub size, and icing conditions with frozen precipitation. Since none of the correlations were serious, all variables were kept in the multivariate model and caution was applied in interpreting the results. This is preferred to the alternative solution of removing variables, which would lead to model misspecification.

Backward stepwise logistic regression was used to calibrate the risk models because of the predictive nature of the research. The selected technique is able to identify relationships missed by forward stepwise logistic regression (Hosmer & Lemeshow 2000, Menard 2001). The predictor variables were entered by blocks, each consisting of related factors, as shown in Table 2, such that the change in the model's substantive significance could be observed as the variables were included. Statistical software SPSS begins by conducting backward stepwise logistic regression on Block 1 variables, removing non-significant variables of that block before conducting backward stepwise logistic regression on the remaining variables from Block 1 and the additional variables from Block 2. This continues until Block 9 variables are included for backward stepwise logistic regression.

Table 2 Blocks of variables

Block	Variables Entered
Block 1	Equipment class, Equipment type
Block 2	User class, Foreign origin/destination
Block 3	Ceiling height
Block 4	Visibility, Fog, Dawn/dusk
Block 5	Crosswind
Block 6	Rain, Electric storm
Block 7	Temperature, Icing conditions, Frozen precipitation, Snow
Block 8	Airport hub size
Block 9	Significant terrain

Table 3 below explains how each variable is measured or categorised for input into the logistic regression. Where entered as categorical variables, the reference category is highlighted in bold.

Table 3 Model Variables Units & Categories

Variable	Variable Type	Categorical Groupings/Measuring Unit
Equipment Class	Categorical	Heavy aircraft – Maximum Take-Off Weight (MTOW)> 255,000lbs Large jets – MTOW between 41,000lbs and 255,000lbs, e.g. Boeing 737, A320 Large commuter aircraft – MTOW between 41,000lbs and 255,000lbs but smaller than large jets, e.g. regional jets Medium aircraft – MTOW between 12,500lbs and 41,000lbs Small aircraft – MTOW under 12,500lbs
Equipment Type	Categorical	Turboprops Jets
User Class	Categorical	Commercial operation Freight operation General aviation operation
Foreign Origin/Destination	Categorical	Domestic Foreign
Ceiling Height	Continuous	100ft
Visibility	Categorical ⁵	<2.00 statute miles (SM) 2.01- 4.00SM 4.01- 6.00SM 6.01- 8.00SM >8.00SM
Fog	Categorical	No Fog Fog
Dawn/Dusk	Categorical	Non dawn/dusk conditions Daen/dusk
Crosswind	Continuous	Knots
Rain	Categorical	No rain Rain

⁵ Visibility was transformed from a continuous variable to a categorical one to meet the logit linearity assumption of logistic regression after Box-Tidwell transformation tests were carried out to ensure that all variables meet this assumption..

Electric Storm	Categorical	No electric storm Electric storm
Temperature	Continuous	10°C
Icing Conditions	Categorical	No icing conditions Icing conditions
Frozen Precipitation	Categorical	No frozen precipitation Frozen precipitation
Snow	Categorical	No snow Snow
Airport Hub Size	Categorical	FAA hub (Large/Medium/Small) FAA non-hub
Significant Terrain	Categorical	No significant terrain Significant terrain ⁶

Cases with missing data were replaced by their respective series means. This only concerned the parameters of ceiling height (47 accidents), crosswind (14 accidents) and temperature (21 accidents). The most severely affected case was ceiling height for take-off overruns, for which 15.4% of accidents had no data and were replaced by the series mean.

With the model coefficients, the probability formula for accident occurrence could be obtained.

For each accident model,

$$P(\text{AccidentOccurrence}) = \frac{1}{1 + e^{-z}}$$

where

$$z = b_0 + b_1(\text{Variable}_1) + b_2(\text{Variable}_2) + \dots + b_n(\text{Variable}_n)$$

where b_0 is the constant and b_1 to b_n are the corresponding parameter coefficients.

Due to the case-control set-up of the study, the constant (intercept) term b_0 of the final formula must be adjusted to account for the different sampling fractions between the

⁶ Defined the same way as in the NOD sampling stratification process

cases and the controls⁷. The following formula was used for this purpose (Hosmer & Lemeshow 2000):

$$b^*_0 = \ln(t_1 / t_0) + b_0$$

where b^*_0 is the original intercept, t_1 is the sampling fraction of cases, t_0 is the sampling fraction of controls and b_0 is the adjusted intercept.

t_1 is one since all relevant accidents have been sampled. From the NOD sampling exercise, it was calculated that the total number of relevant normal operations from 2000 to 2005 inclusive is 191,902,290 operations. That is 44.78 percent of the period's total itinerant operations excluding military operations. From the TAFs, the total number of itinerant operations from 1982 to 2002 inclusive (the accident sampling period) excluding military operations was computed to be 1,408,495,828 movements. 44.78 percent of the latter equates 630,792,133 movements. Since the total sampled normal operation population is 242,420 flights,

$$\begin{aligned} t_0 &= 242420/630792133 \\ &= 3.843 \times 10^{-4} \end{aligned}$$

With t_1 and t_0 , the adjusted intercepts of each of the risk model formula could be calculated:

$$b^*_0 = \ln(t_1 / t_0) + b_0 = \ln(1/3.843 \times 10^{-4}) + b_0 = 7.864 + b_0$$

Table 4 shows the original and adjusted intercepts.

Table 4 Original & adjusted risk model equation intercepts

Model	Original intercept	Adjusted intercept
LDOR	-8.431	-16.295
LDUS	-8.911	-16.775
TOOR	-9.281	-17.145
TOC	-9.540	-17.404

⁷ If left unadjusted, the intercept of the risk model formula would suggest that the sampling fraction of cases and controls are identical and so yield unrealistically high accident probabilities given actual difference in sampling fractions.

With the adjusted intercept term, the z for the landing overrun probability formula is:

$$\begin{aligned}
 z = & -16.295 + 0.486(\text{HeavyAcft}) - 1.631(\text{L arg eCommuterAcft}) + 0.893(\text{MediumAcft}) + 1.951(\text{SmallAcft}) \\
 & + 1.050(\text{TurbopropAcft}) + 0.934(\text{FreightOp}) + 0.835(\text{GAOp}) - 1.565(\text{ForeignOD}) - 0.014(\text{CeilingHeight00 ft}) \\
 & + 1.443(\text{Visibility} < 2SM) - 0.239(\text{Visibility}2 - 4SM) - 1.429(\text{Visibility}4 - 6SM) + 0.276(\text{Visibility}6 - 8SM) \\
 & + 2.437(\text{Fog}) + 0.486(\text{DawnDusk}) + 0.089(\text{Crosswindknts}) + 2.164(\text{IcingConditions}) + 1.860(\text{Snow}) \\
 & + 0.588(\text{NonhubApt}) + 0.417(\text{SignificantTerrain})
 \end{aligned}$$

The z for the landing undershoot probability formula is:

$$\begin{aligned}
 z = & -16.775 + 0.139(\text{HeavyAcft}) - 2.017(\text{L arg eCommuterAcft}) + 1.457(\text{MediumAcft}) + 2.932(\text{SmallAcft}) \\
 & + 1.086(\text{TurbopropAcft}) + 0.894(\text{FreightOp}) + 0.610(\text{GAOp}) - 0.017(\text{CeilingHeight00 ft}) + 1.881(\text{Visibility} < 2SM) \\
 & + 0.446(\text{Visibility}2 - 4SM) - 0.234(\text{Visibility}4 - 6SM) + 0.321(\text{Visibility}6 - 8SM) + 1.738(\text{Fog}) + 0.043(\text{Crosswindknts}) \\
 & + 3.775(\text{IcingConditions}) - 2.562(\text{Frozen Precipitation}) + 2.011(\text{Snow}) + 0.819(\text{SignificantTerrain})
 \end{aligned}$$

The z for the take-odd overruns probability formula is:

$$\begin{aligned}
 z = & 17.145 + 1.157(\text{HeavyAcft}) - 0.485(\text{L arg eCommuterAcft}) + 2.082(\text{MediumAcft}) + 3.860(\text{SmallAcft}) \\
 & + 0.968(\text{ForeignOD}) - 0.008(\text{CeilingHeight00 ft}) + 0.320(\text{Visibility} < 2SM) - 2.077(\text{Visibility}2 - 4SM) \\
 & - 0.470(\text{Visibility}4 - 6SM) - 0.544(\text{Visibility}6 - 8SM) + 1.847(\text{Fog}) + 0.093(\text{Crosswindknts}) \\
 & - 0.254(\text{Temperature}) + 2.932(\text{Snow})
 \end{aligned}$$

The z for the crashes after take-off probability formula is:

$$\begin{aligned}
 z = & -17.404 + 0.760(\text{HeavyAcft}) - 0.776(\text{L arg eCommuterAcft}) + 1.251(\text{MediumAcft}) + 2.842(\text{SmallAcft}) \\
 & + 0.934(\text{TurbopropAcft}) + 2.049(\text{FreightOp}) + 1.316(\text{GAOp}) - 0.003(\text{CeilingHeight00 ft}) + 1.307(\text{Visibility} < 2SM) \\
 & - 0.790(\text{Visibility}2 - 4SM) - 1.104(\text{Visibility}4 - 6SM) + 0.178(\text{Visibility}6 - 8SM) + 1.753(\text{Fog}) + 0.683(\text{DawnDusk}) \\
 & + 0.074(\text{Crosswindknts}) + 2.246(\text{IcingConditions}) - 2.188(\text{Frozen Precipitation}) + 2.561(\text{Snow}) - 0.734(\text{NonhubApt}) \\
 & - 1.213(\text{SignificantTerrain})
 \end{aligned}$$

Using the formulae above and the measurement details from Table 3, the accident probability of specific flights can be assessed. For categorical variables, only multiply the relevant coefficient by 1 and multiply the irrelevant categories by 0.

It can be seen that the four formulae do not contain identical parameters. The stepwise regression procedure has eliminated parameters that are not significant for the particular risk models. All remaining parameters are significant at the 95 percent

confidence interval. For example, foreign origin/destination only features in the formula for landing overruns and take-off overruns. Moreover, their signs are also different. Foreign operation is negative in the landing overrun formulae and positive in the take-off overrun one. This indicates that the factor contributes to accident risk for take-off overruns but has the opposite effect on landing overruns. Indeed, only 1.0 percent of landing overruns involve a foreign origin or destination whereas 11.5 percent of take-off overruns do. For landing undershoots and crashes after take-off, the variable was removed by the stepwise regression as a significant variable even before the next variable was entered. This is likely to be related to the strong explanatory power of preceding variables (equipment class, type and user class). The great majority of factors, however, bear the same sign for all accident types. The size of the factors' coefficients also differs between the four formulae. The coefficients for fog, for instance, vary from 1.738 (landing undershoot) to 2.437 (landing overrun). This reflects the degree to which the factor increases accident risk.

6. Model Goodness-of-Fit

To assess the models' goodness-of-fit, the Nagelkerke R^2 measures of the respective models were calculated and shown in Table 5⁸.

Table 5 Model Nagelkerke R^2

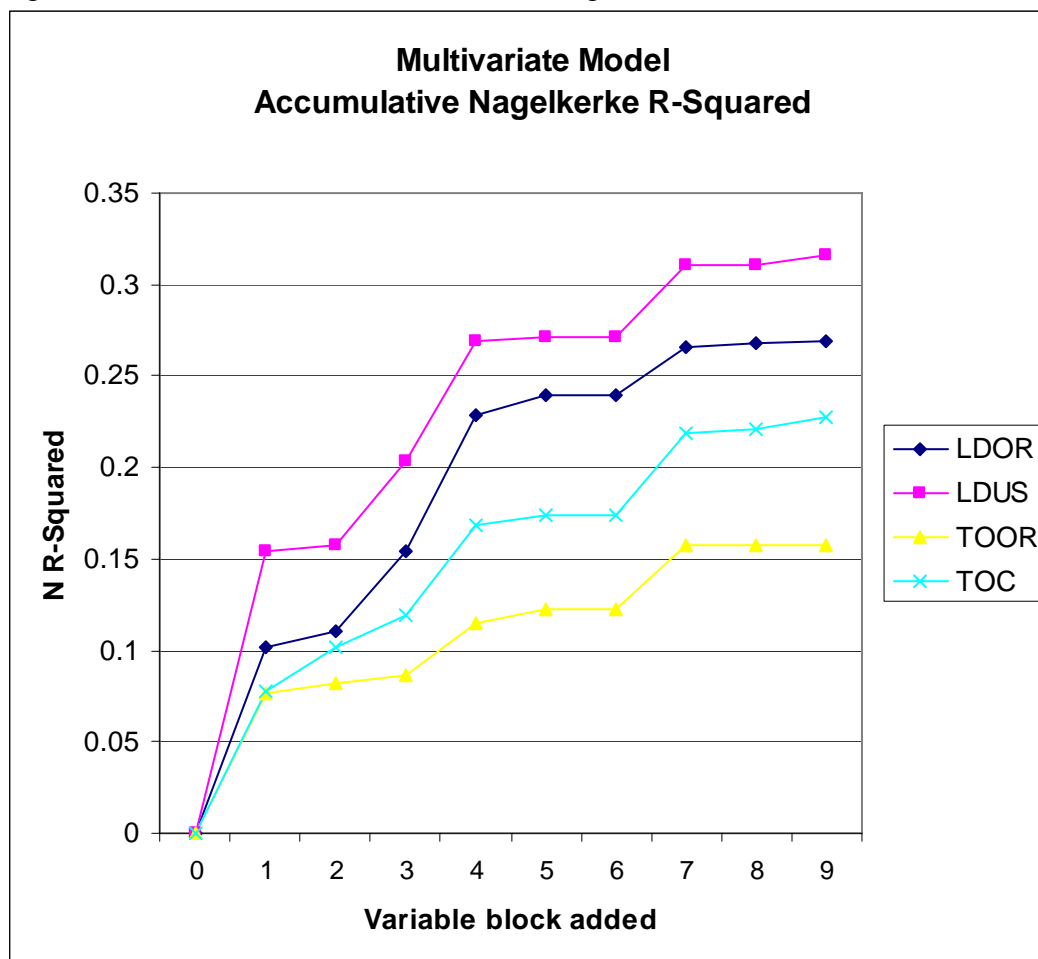
Model	Nagelkerke R^2
Landing Overrun	0.269
Landing Undershoot	0.316
Take-off Overrun	0.157
Crash after Take-off	0.227

The model for landing undershoot occurrence is the most potent, explaining twice as much data variation than the model for take-off overruns, the worse-performing model. Relatively low R^2 values are the norm in logistic regression (Ash & Schwartz 1999) and they should not be compared with the R^2 of linear regressions (Hosmer &

⁸ Nagelkerke R^2 is a pseudo measure of model substantive significance similar to R^2 in linear regression, varying between zero and one. Nagelkerke R^2 of 0.3 suggests that the model explains roughly 30 percent of the variance in the data.

Lemeshow 2000). Figure 2 shows how Nagelkerke R^2 increased as variables were added to the model.

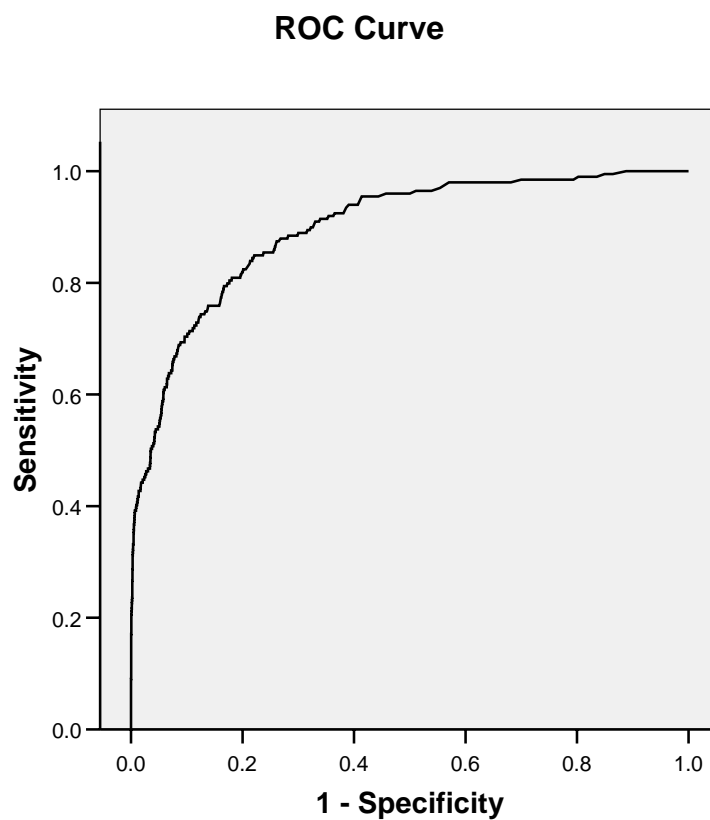
Figure 2 Multivariate model accumulative Nagelkerke R^2



In order to assess how successful the models are in classifying flights correctly as “accident” or “normal” and to find the appropriate cut-off points for the logistic regression models, Receiver Operating Characteristics (ROC) Curves were used. The cut-off point is the critical probability above which the model will class an event as an accident. The ROC curve plots all potential cut-off points according to their respective True Positive Rates (percentage of accidents correctly classed as accidents) and False Positive Rates (percentage of normal flights incorrectly classed as

accidents). The best cut-off point would have a optimally high TPR and low FPR. Figures 3 to 6 display the four models' ROC curves. TPR is labelled Sensitivity and FPR 1-Specificity.

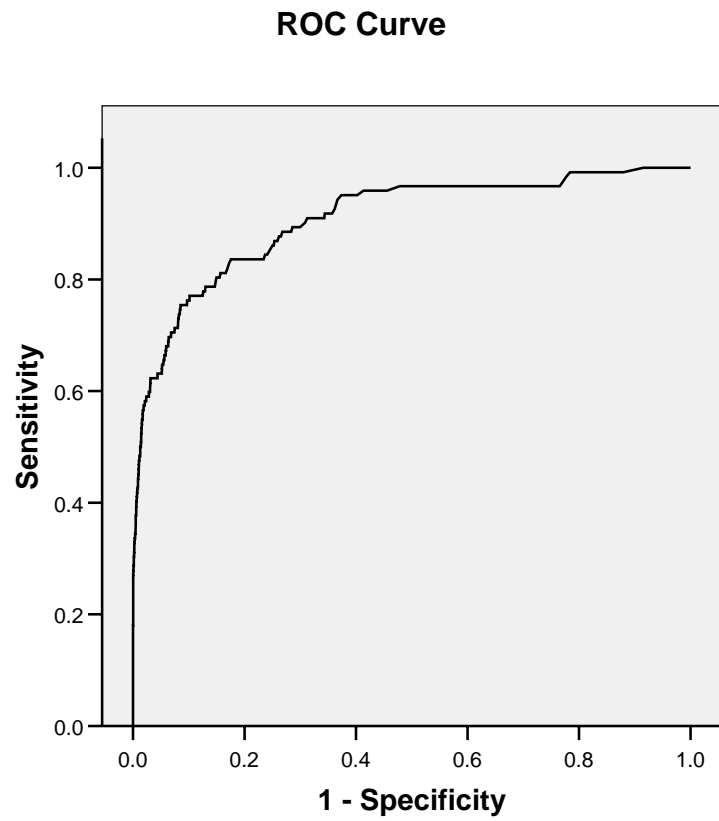
Figure 3 Landing overrun model ROC curve



Diagonal segments are produced by ties.

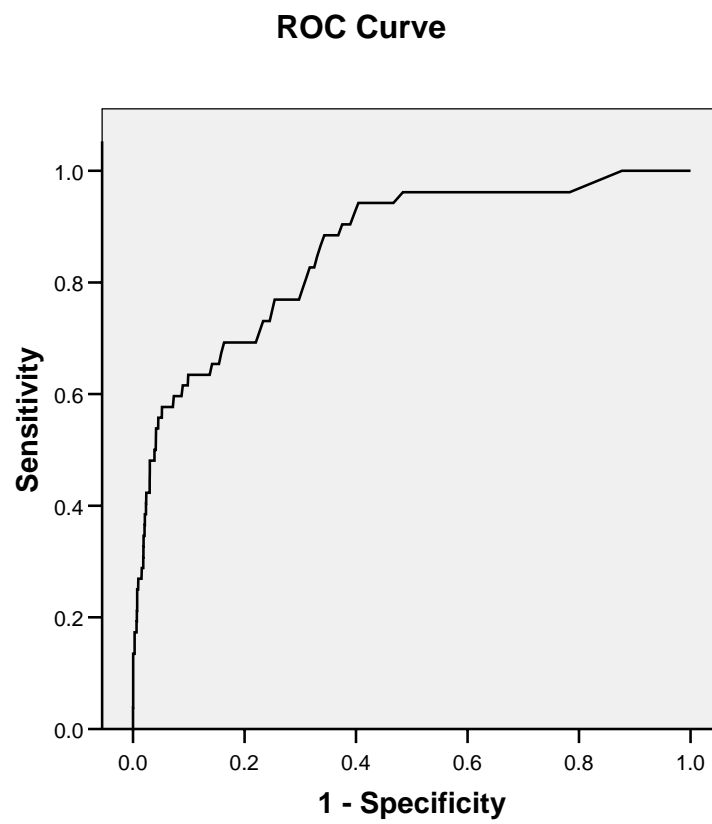
The ROC curve graphically presents the trade-off between TPR and FPR for all possible cut-off points, the best of which is likely to be the point closest to the top-left corner of the graph. The trade-off between TPR and FPR can be seen in Figure 11.2. As the TPR (sensitivity) rises, the FPR (1-specificity) also increases. The larger the area under the curve, the better the model is at identifying accidents from normal flights. Figures 4 to 6 are interpreted in the same way. It is clear that the landing accident models produced better results than the take-off models.

Figure 4 Landing undershoot model ROC curve



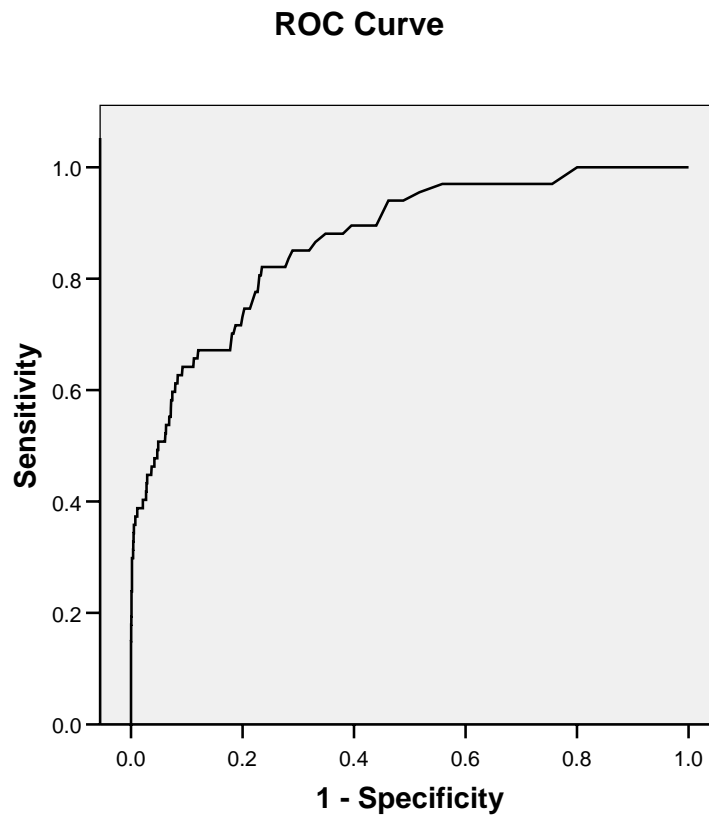
Diagonal segments are produced by ties.

Figure 5 Take-off overrun model ROC curve



Diagonal segments are produced by ties.

Figure 6 Crash after take-off model ROC curve



Diagonal segments are produced by ties.

The area under the ROC curve is quantified by the c statistic, which measures the discriminative power of the accident frequency models. The statistic varies between 0.5 (indicating that the model's predictions are no better than chance) and 1 (indicating a perfect classification model with 100 percent TPR and 0 percent FPR). Table 6 shows the c statistics for the four models.

Table 6 Model c statistics

Model	c statistic
Landing Overrun	0.897
Landing Undershoot	0.908
Take-off Overrun	0.858
Crash after Take-off	0.868

c statistics of over 0.9 suggest excellent classification accuracy and ones above 0.8 are considered good (Tape 2007). Therefore, all models have at least good classification accuracy. As expected, the order of the c statistic findings reflects those of the Nagelkerke R^2 in Table 2. The models' performance seem significantly better when measured by c statistics than Nagelkerke R^2 because the former is not dependent on the frequency of the outcome, whereas R^2 is smaller when the outcome is infrequent, which is true for accident occurrence (Ash & Schwartz 1999).

Table 7 further compares the TPR and FPR of the four models at selected cut-off points.

Table 7 True positive rate & false positive rate comparison

Model	Cut-off point	TPR	FPR
Landing Overrun	0.00000052127	0.849	0.221
Landing Undershoot	0.00000024800	0.844	0.237
Take-off Overrun	0.00000010955	0.846	0.330
Crash after Take-off	0.00000009420	0.851	0.290

For the landing overrun model, then, a cut-off point of 0.00000052127 yields 84.9 percent of accidents being correctly classed as such and 22.1 percent of normal flights falsely classified as accidents. The latter could in fact be interpreted as high risk but incident-free operations. Depending on the objective of the risk assessment exercise, a relatively conservative or risk-tolerant cut-off point could be chosen.

7. Comparison with Standard Risk Assessment Models

Most standard risk assessments for airport safety areas rely on simple crash rates according to general groupings of aircraft type when considering accident frequency.

It is evident from Figure 1 that the final models' substantive significance as measured by Nagelkerke R^2 are considerable improvements upon models based only on Block 1 parameters (aircraft size and type). The results also compare favourably with Kirkland's landing overrun model based on excess runway distance, which only explained 11 percent of risk determinants (Kirkland 2001).

Figure 7 shows the difference in terms of c statistics between standard models and the improved ones. The former are defined as models that only consider aircraft size and engine type. Gains in goodness-of-fit and predictive power were observed for models of all accident types.

Figure 7 c statistic comparison

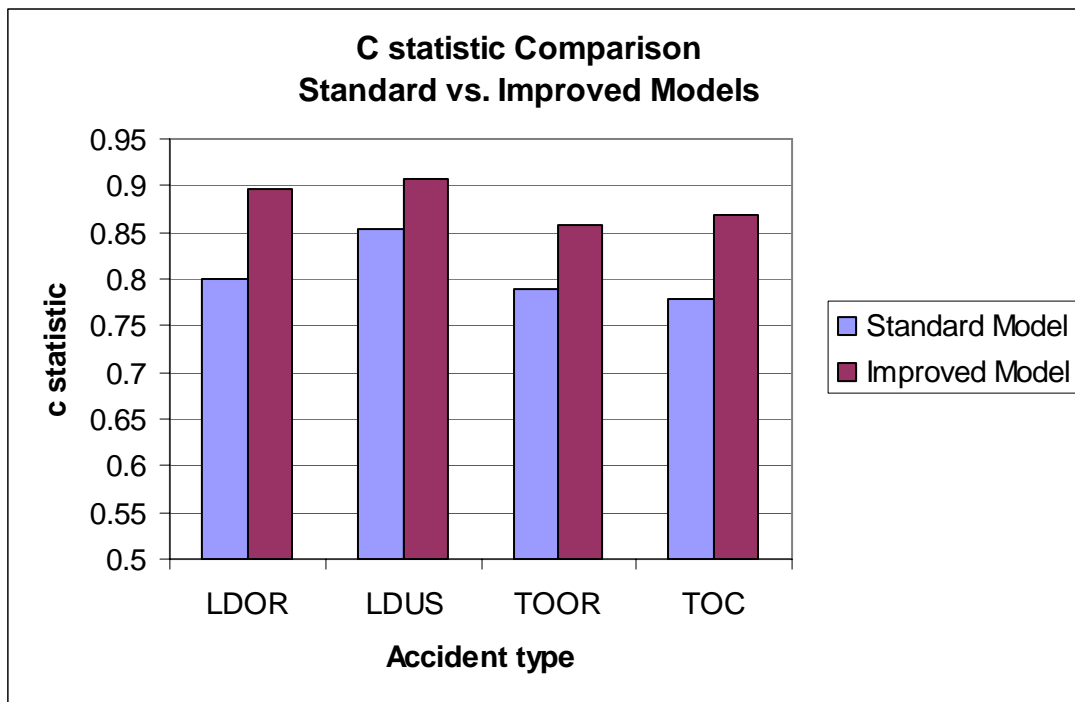


Table 8 then contrasts the models' predictive accuracy by comparing their respective false positive rates at cut-off points with similar true positive rates.

Table 8 False positive rate comparison

Model	Standard Model TPR	Improved Model TPR	Standard Model FPR	Improved Model FPR
Landing Overrun	0.844	0.849	0.395	0.221
Landing Undershoot	0.852	0.844	0.299	0.237
Take-off Overrun	0.846	0.846	0.405	0.330
Crash after Take-off	0.821	0.851	0.405	0.290

At similar TPRs, the FPR of the improved models are significantly lower than that of the standard models for all accident types. For example, at a true positive rate of 84.6 percent, the model of take-off overruns using the conventional predictor variables incorrectly classed 40.5 percent of normal operations as accidents. In contrast, the equivalent for the model developed in the present study is 33.0 percent. The models' increased ability in discriminating between safe and accident flights are important steps towards better airport risk assessment. While there is certainly still room for improvement, from the various measures, it is nonetheless clear that important gains have been made in improving the accident frequency models' fit and predictive power for better airport risk assessment. The model could additionally be used for other related purposes, such as a risk assessment tool for pilots before take-off and landings as well as for designing approach procedures adapted to different landing conditions.

8. Conclusion

Using a single comprehensive database of relevant accidents, multidimensional NOD and multivariate modelling, improved accident frequency models were developed, addressing many of the fundamental deficiencies of standard airport risk assessment methodologies. The current study has taken an integrated approach in assessing all accident types related to airport safety areas and has built a single comprehensive accident database accordingly. A large Normal Operations database was also developed such that risk exposure data could be incorporated into the modelling exercise to quantify the criticality of risk factors, hence contributing to a more sensitive risk assessment technique. The scope of the accident and Normal Operations databases allow an unprecedented number of risk factors to be included in the risk models, encompassing aircraft, operational and notably a series of meteorological factors.

Using multivariate logistic regression, frequency models for the four relevant accident types were calibrated. Where normal operations data was used before (Enders et al. 1996, Kirkland 2001), it was one-dimensional in nature, which limits modelling capability and fails to account for joint influences between variables. In contrast, the multidimensional model developed adjusts for the joint influences between risk factors and provides a single risk formula for the combined effects of multiple risk factors. It is able to offer risk estimates for individual flights as well as assess the risk profile of an airport with specific traffic and environmental characteristics. The increased modelling capability and flexibility add much value to the models as risk assessment tools.

Multivariate modelling using the range of risk factors available improved predictive power compared to previous methodologies, including Kirkland et al. (2003), and standard methods that only considered aircraft and engine types. Nagelkerke R^2 and the c statistic were used to assess the goodness-of-fit and predictive ability of the models. On average, the models developed in this study explained 14 percentage points more data variation than conventional models. Standard techniques that only included aircraft and engine type had an average c statistic of 0.81 across the accident types, whereas the improved models averaged 0.88. Improvements in model sensitivity and specificity were also observed. The final part of this two-part paper demonstrates the use of the improved models through case studies of two airports.

References

- Ash, A., Schwartz, M. 1999. R^2 : a useful measure of model performance when predicting a dichotomous outcome. *Statistics in Medicine*, February 28; 18(4), p.375-384
- Button, K., Drexler, J., 2006. Are measures of air-misses a useful guide to air transport safety policy? *Journal of Air Transport Management*. Vol. 12, p.168-174
- CAA, 1998. Risks from aeroplanes overrunning aerodrome runways. Edition one. Civil Aviation Authority, London
- DfT, 1997. Third party risk near airports and Public Safety Zone policy. Department for Transport, London
- DOT, 1979. Development of a systems risk methodology for single and multi-modal transportation system. Department of Transport, Washington, DC
- Eddowes, M., Hancox, J., MacInnes, A., 2001. Final report on the risk analysis in support of aerodrome design rules. Report for the Norwegian Civil Aviation Authority. AEA Technologies plc, Warrington, UK
- Enders, J., Dodd, R., Tarrel, R., Khatwa, R., Roelen, A., Karwal, A., 1996. Airport Safety: A study of accidents and available approach-and-landing aids. *Flight Safety Digest*, March. Flight Safety Foundation, VA, USA
- FAA, 2007. Introduction, 7th Instrument Approach Procedure. US Terminal procedure Publications. Federal Aviation Administration, Washington DC
http://avn.faa.gov/content/naco/online/pdf_files/7th_iap_intro.pdf, accessed 27/02/07
- Hale, A. 2001. Regulating airport safety: the case of Schiphol. *Safety Science*. Vol.37, p.127-149
- Hale, A., 2002. Risk contours and risk management criteria for safety at major airports, with particular reference to the case of Schiphol. *Safety Science*. Vol. 40, p.299-323
- Helmreich, R., Klinect, J., Wilhelm, J., 1999. Models of threat, error, and CRM in flight operations. *Proceedings of the 10th International Symposium on Aviation Psychology*, Columbus, p.677
- Hosmer, D., Lemeshow, S., 2000. *Applied logistic regression*. John Wiley & Sons, Hoboken
- ICAO, 2006. *Safety Management Manual*. First edition. Doc. 9859. International Civil Aviation Organisation, Montreal

- Khatwa, R., Roelen, A., Karwal, A., Enders, J., Dodd, R., Tarrel, R., 1996. An evaluation of approach landing factors influencing airport safety. NLR-TP-96221U. Dutch National Aerospace Laboratory NLR, Amsterdam
- Khatwa, R., Helmreich, R., 1998. Analysis of critical factors during approach and landing in accidents and normal flight. Final report version 2.0. Flight Safety Digest. Vol. November 1998-February 1999. Flight Safety Foundation, VA, USA
- Kirkland, I., 2001. The risk assessment of aircraft runway overruns accidents and incidents. Doctoral thesis. Loughborough University, Loughborough, UK
- Kirkland, I., Caves, R., Pitfield, D., 2003. An improved methodology for assessing risk in aircraft operations at airports, applied to runway overruns. Safety Science. Vol. 42, p.891-905
- Klinec, J., Wilhelm, J., Helmreich, R., 1999. Threat and error management: Data from line operations safety audits. Proceedings of the 10th International Symposium on Aviation Psychology, Columbus, p.683
- Li, G., Baker, S., Grabowski, J., Rebok, G., 2001. Factors associated with pilot error in aviation crashes. Aviation, Space, and Environmental Medicine. Vol. 72, p.52-58
- Menard, S., 2001. Applied logistic regression analysis. 2nd edition. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-103. Sage, Thousand Oaks
- Myers, R., 1990. Classical and modern regression with applications. 2nd edition. Duxbury, Boston
- Nagelkerke, N., Smits, J., Le Cessie, S., van Houwelingen, H., 2005. Testing the goodness-of-fit of the logistic regression model in case-control studies using sample reweighting. Statistics in Medicine. Vol. 24, p.121-130
- Piers, M., Loog, M., Giesberts, M., Moek, G., Couwenberg, M., Smeets, M., 1993. The development of a method for the analysis of societal and individual risk due to aircraft accidents in the vicinity of airports. NLR-CR-93372L. Dutch National Aerospace Laboratory NLR, Amsterdam
- Piers, M., 1994. The development and application of a method for the assessment of third party risk due to aircraft accidents in the vicinity of airports. 19th International Council of Aeronautical Sciences Congress, Anaheim September ICAS proceedings. ICAS, Vol. 19, p.507-518

- Piers, M., 1998. Methods and models for the assessment of third party risk due to aircraft accidents in the vicinity of airports and their implications for societal risk. *Technology Risk and Society*. Vol. 12, p.166-204
- Roelen, A., Bellamy, L., Hale, A., Molemaker, R., van Passen, M., 2000. Feasibility of the development of a causal model for the assessment of third party risk around airports. Part 1: Main Report. NLR-CR-2000-189-PT-1. Dutch National Aerospace Laboratory NLR, Amsterdam
- Roelen, A., Wever, R., Hale, A., Goossens, L., Cooke, R., Lopuhaa, R., Simons, M., Valk, P., 2002. Causal modelling of air safety. Demonstration model. NLR-CR-2002-662. Dutch National Aerospace Laboratory NLR, Amsterdam
- Tape, T., 2007. Interpreting diagnostic tests. University of Nebraska Medical Center. <http://gim.unmc.edu/dxtests/roc3.htm>, accessed 25/01/07
- Van Es, G., Van der Geest, P., Nieuwpoort, T., 2001. Safety aspects of aircraft operations in crosswind. NLR-TP-2001-217. Dutch National Aerospace Laboratory NLR, Amsterdam
- Watson, D., 2005. Summary and analysis of, and CAA response to, comments and submissions on NPRM 04-03 received during public consultation. Civil Aviation Authority of New Zealand, Petone
- Wong, D., Pitfield, D., Caves, R., Appleyard, A., 2006. The development of aircraft accident frequency models. In: Guedes Soares & Zio. *Safety and Reliability for Managing Risk. Proceedings of the European Safety and Reliability Conference 2006 (ESREL 2006)*, 18-22 September, Estoril. Taylor & Francis, London, Vol. 1, p.83-90
- Wong, D., 2007. The modelling of accident frequency using risk exposure data for the assessment of airport safety areas. PhD thesis, Loughborough University