Australian Government
Civil Aviation Safety Authority
www.casa.gov.au

PETA – Review of Current Trend Analysis Practices within CASA

Publishing information

Published by:	Civil Aviation Safety Authority
Postal address:	GPO Box 2005, Canberra ACT 2601
Office:	16 Furzer Street, Phillip ACT 2606
Telephone:	131 757 +61 2 6217 1111 (from outside Australia)
Facsimile:	+61 7 3144 7575
Email:	safetysystems@casa.gov.au
Internet:	https://www.casa.gov.au/

© Civil Aviation Safety Authority 2015

Ownership of intellectual property rights in this publication

Unless otherwise noted, copyright (and any other intellectual property rights, if any) in this publication is owned by the Civil Aviation Safety Authority (referred to below as CASA).

Disclaimer

The material contained in this publication is made available on the understanding that CASA is not providing professional advice, and that users exercise their own skill and care with respect to its use, and seek independent advice if necessary.

CASA makes no representations or warranties as to the contents or accuracy of the information contained in this publication. To the extent permitted by law, CASA disclaims liability to any person or organisation in respect of anything done, or omitted to be done, in reliance upon information contained in this publication.

Creative Commons licence

With the exception of the Coat of Arms, CASA logo, and photos and graphics in which a third party holds copyright, copyright in this publication is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License is a standard form licence agreement that allows you to copy and redistribute the material in any medium or format provided that you attribute the work to CASA and abide by the other licence terms.

A summary of the licence terms is available from:
http://creativecommons.org/licenses/by-nc-nd/4.0/

The full licence terms are available from:
http://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.

DATA

STATISTICS

ANALYSIS

MATHEMATICS

AUSTRALIA

INFORMATION

Review of Current Trend Analysis Practices within CASA

July 2013

Project CASA/1

STRATEGIC
INFORMATION
CONSULTANTS

Review of Current Trend Analysis Practices within CASA

September 2013

Client: Civil Aviation Safety Authority

Project: CASA/1

Consultants: Dr Kathy Haskard
Dr John Henstridge
Yuichi Yano
Dr Alethea Rea

Data Analysis Australia Pty Ltd
97 Broadway
Nedlands, Western Australia 6009
(PO Box 3258 Broadway, Nedlands 6009)
Website: www.daa.com.au
Phone: (08) 9386 3304
Facsimile: (08) 9386 3202
Email: daa@daa.com.au
A.C.N. 009 304 956
A.B.N. 68 009 304 956

Executive Summary

Data Analysis Australia has reviewed the statistical methods currently used by the Safety Performance Analysis (SPA) Branch of the Civil Aviation Safety Authority (CASA) for screening aviation safety incident data to highlight incidents or issues requiring further investigation by CASA. The review considered the Systematic Trends and Analysis Reporting Tool (START), a Microsoft Excel–based tool developed by CASA, and the similar calculations currently performed in an Access database via SQL. In carrying out the review, Data Analysis Australia considered best modern statistical practice while at the same time recognising that solutions needed to be practical and commensurate with the screening nature of the tool.

The safety analysis procedure currently employed by CASA embodies several excellent characteristics, particularly the focus on unique occurrences and the fitting of a trend model to the historical data. At the same time, the implementation neglects a number of statistical principles; in particular:

- The linear model fitted to historical data implicitly assumes that the historical data is normally distributed, when count data cannot be. This is particularly serious when counts are small as is the case for many types of occurrences. We believe it is **essential** to move to a Poisson distribution–based method which will better reflect the nature of the data.

- The method of identifying whether the last observed data value – the count of occurrences for the last quarter – is unexpectedly high or low is done by comparison with a trend model including that value. This loses statistical power and can lead to anomalies. We **recommend** that the comparison be made to a trend model excluding the last data value, and suggest methods by which this can be readily done in the Poisson regression context.

- There is one way that conditions remain the same, but many ways in which they can change. The Poisson models can be used to understand this change. We **recommend** that the analysis be extended to highlight the types of change observed.

- The measures of high variability currently used are simple measures of total variability that do not distinguish between variability due to trend, small numbers of occurrences, or true instability. We **recommend** improved measures related to the Poisson model that separate out the above components of variation.

- The current computational framework in SQL is working at its practical limits and, while automating the process, is actually a simplification of the methods used in the Excel tool. It is clear that a more appropriate computational framework is required. We **recommend** using the statistical package R and having this automatically called from either Access (or other SQL system) or Excel.

- We **encourage** CASA to pursue acquisition of data on exposure to risk of incidents, where this might have changed over time. This can be incorporated

into the statistical model, enabling changes in the *rate* of incidents per unit exposure to be modelled and investigated.

The above improvements are believed to be both best practice and very practical. Data Analysis Australia suggests a tabular display of the outcomes, with one line for each occurrence type, summarising the flags raised for different aspects in the data. This table would be supplemented by graphical plots of the data overlaid with the most relevant fitted models for each occurrence type. These are an invaluable aid to visualisation, interpretation and further investigation. This approach reflects the outputs from the improved analysis in a compact way while allowing the new statistical procedures to be introduced within CASA's existing general philosophy.

Data Analysis Australia will prepare an implementation plan for the recommended Poisson regression approach, to be conducted in the public domain statistical software R, with a possible interface to an Excel spreadsheet.

Table of Contents

1. Introduction

The Safety Performance Analysis (SPA) Branch of the Civil Aviation Safety Authority (CASA) is responsible for conducting analysis and reporting of safety-related trends and risk factors to CASA's Executive Management. CASA has developed a trend analysis tool that is currently used for monitoring and reporting relevant developments in the safety-related incident data.

CASA is considering a further development of the capability of their trend analysis through integration of an independent trend analysis, which allows more complex analysis to be undertaken, as part of the current trend analysis practice. The review of current trend analysis practices within CASA is deemed a necessary phase in this development process.

This report outlines the review of the current analysis and reporting of safety trends.

1.1 Scope of Work

Data Analysis Australia has carried out the review noting that the four key areas of interest to CASA are:

- An assessment of current methodology: Is this methodology acceptable to use on the datasets to which CASA has access? Are there other suitable methods that could be used with this data?
- Limitations of the current methodology: To what extent can very low counts be reported on? What level of confidence can be placed in the methodology?
- Assessment of appropriate aggregation period of the data: Should the aggregation continue to be quarterly? Would monthly or annually be more appropriate? Should seasonality be analysed and reported?
- Recommendation of additional statistics to be included in the report – analysis not currently conducted and reported that would be valuable to consider in future.

Throughout the report Data Analysis Australia has focussed on these four areas, particularly in the section on specific recommendations (Section 4).

2. Current Methodology used by CASA

2.1 Description of Data

This section outlines the sources of data available to CASA and some of the key features of the data.

Some of the key features include the fact that safety incident data, by its very nature, is count data. Furthermore it is likely that these counts will be low for some groups (like occurrence types for severe incidents). Each occurrence type (and indeed each occurrence) will have an associated severity. This severity may be difficult to quantify. There is also potential for incidents to occur in clusters (say poor maintenance at a particular location).

Another feature of safety incident data is that underreporting may have the same features as an improved safety record and as such it may be difficult to say with any certainty that safety has improved over time.

It also should be noted that much of the information on incidents is qualitative and therefore analysis of the quantitative components is only a part of any process of investigation into safety trends.

CASA currently has access to data from numerous sources, some owned by external agencies and some maintained by departments within CASA. These sources are listed in Table 1. CASA also actively seeks out other sources of information to incorporate in analyses, including industry information (for example information from the International Air Transport Association (IATA) and the Flight Safety Foundation).

The main data source used by CASA is the Aviation Safety Incident Reports (ASIR) sourced from the Australian Transport Safety Bureau (ATSB), with other sources providing supplementary information in reporting different categories of the safety incidents.

CASA expects future changes to the data sourcing procedure, where some data sources, such as Electronic Safety Incident Reports (ESIR), will change with the introduction of a new system in Airservices Australia. Some fields currently recorded by the ESIR dataset will not be reported to the ATSB. There is also a prospect of obtaining new data sources to improve the chances of detecting trends in incidents.

Data Source	Owner
Aircraft Register	
AOC Survey Data (AHSQ)	Safety Systems Office
Aviation Safety Incident Report (ASIR)	Australian Transport Safety Bureau (ATSB)
Electronic Safety Incident Report (ESIR)	Airservices Australia
Industry Contact Register	Safety Education and Promotion(Education)
iStars	ICAO
Service Difficulty Reports (SDRs)	Airworthiness and Engineering Branch
Surveillance Findings (Sky Sentinel)	Operations Division

Table 1. Data sources available to CASA in reporting Aviation Safety Incidents.

The example Safety Performance Analysis Quarterly Analysis Report provided to Data Analysis Australia used only two data sources, the Aviation Safety Incident Reports (ASIRs) and Electronic Safety Incident Reports (ESIRs), both of which relate to industry occurrences.

The ASIRs, maintained by the ATSB, have an occurrence ID, category type, date and time, location (including the latitude and longitude), the occurrence type descriptions (at three levels), information on the aircraft itself, the phase of flight when the incident occurred and the aircraft airspace class type. The ASIRs contain non-standard values for aircraft manufacturer and model, so CASA manually maps this information before analysis. The ASIRs also contain a summary field, which describes the incident. A table of fields in the ASIRs, as seen in the extract provided, is given in Table 2.

The ASIR data can have more than one record for each occurrence ID. This can arise if more than one aircraft is involved or if the incident can be assigned to multiple occurrence types. Therefore these linked incidents require careful consideration in reporting, and can in themselves highlight clusters of events. If the linked incidents are of different occurrence types then such a cluster will contribute to several reports of occurrence type but probably only one location meaning clusters are still likely to be detected.

Because the ESIR are used as supplementary data and the majority of the key information is contained in ASIR, the ESIR data was not reviewed and its details will not be discussed in this report. However, the implication of the existence of overlapping safety incidents reported in ESIRs (and possibly in the reports from other sources) and ASIRs has been noted, and SPA confirmed that the data from the two sources is used independently. The ESIR continues to be used for the purpose of reporting incidents by operator.

Fields in ASIR	Notes
Occurrence ID	Numeric
Occurrence ASIR Summary	
Occurrence Category Type	Accident/Incident/Serious Incident
Occurrence Date and Time	dd/mm/yyyy HH:MM
Occurrence Date	Same format as the Date and Time field but with time set to 0:00
Occurrence Location (Cleaned)	
Occurrence Longitude (Decimal Degrees)	
Occurrence Latitude (Decimal Degrees)	
Occurrence Type Is Primary (Flag)	Y/N
Occurrence Type Description Level 1	
Occurrence Type Description Level 2	
Occurrence Type Description Level 3	May be blank
Aircraft Type	May be blank
Aircraft Engine Type	NA if Aircraft Type is blank
Aircraft Model Common Name	NA if Aircraft Type is blank
Aircraft Operation Sub Type	NA if Aircraft Type is blank
Aircraft Operation Type	NA if Aircraft Type is blank
Aircraft ICAO Type Designator	
Aircraft Phase Of Flight Type	NA if Aircraft Type is blank
Aircraft Airspace Class Type	NA if Aircraft Type is blank

Table 2. Data fields in ASIR.

There have been some changes in taxonomy over time, and these require particular attention. For example, prior to July 2009 "Landing gear unsafe indication" and

"Landing gear" were recorded separately. For the remainder of 2009 these two categories were used, as was a new category "Landing gear/Indicator On". For analysis these three classes should either be grouped or analysis be carried out from February 2010 onwards when a single category captures all the incidents. Typically, when the ATSB makes taxonomy changes it also backcasts the series and informs CASA.

2.2 Particular Features of Aviation Incident Data

- Count data – each incident is a discrete event.
- Many incident types have a severity.
- Potential for clusters of incidents.
- Missing data would look the same as no incidents.
- Much information might be qualitative.

2.3 Description of the Current Methodology

Currently CASA produces quarterly and ad hoc reports based on the Systematic Trends and Analysis Reporting Tool (START). The quarterly reports have two main components, trend monitoring for abnormalities and a summary of ATSB findings regarding trends in ASIR occurrence types. Data Analysis Australia is reviewing the trend monitoring.

The summary sheet of the quarterly report gives the actual and expected number of incidents. The findings from the ESIRs and ASIRs are reported separately. The data provided from the ESIRs is used to report on operators and occurrence types while the data from the ASIRs is used to report on location, aircraft type, occurrence type, and state.

There are five broad categories of occurrence types, namely operational; mechanical; airspace; aerodromes and airways facility; and environment, but the occurrences are reported at a finer level, by their level three occurrence type.

2.3.1 Reporting of Incidents

ASIRs are reported by industry with review and classification of information undertaken by the data owner (the ATSB). CASA is not involved in the classification and review of the occurrence data supplied by Airservices Australia and the ATSB. The delay between an occurrence and reporting of the incident varies for each incident. The need for investigation of an incident is assessed based on criteria such as, but not limited to, those dictated in the ESIR Business Rules (Effective as at 21 December 2011):

- Potential safety value gained from the investigation;
- Likelihood of recurrence and associated worst credible consequence;
- Potential for discovering systemic human and procedural errors and equipment defects;

- Frequency of the occurrence where it suggests a previously unexplored or unmanaged safety issue;
- Profile of the occurrence – in particular the potential impact on confidence in the safety of Airservices' operations;
- Legal or regulatory requirement;
- Available resources.

It is expected that majority (approximately 80%) of the incidents are reported within a few days of the occurrence, and that the delay should not exceed 6 weeks for the longest cases. Given CASA reports the safety incidents on quarterly basis, with a one quarter lag, most of the incidents are expected to be included in the report. ASIR information provided to CASA is updated weekly via an automated feed.

2.3.2 Data Preparation

CASA has been developing its trend analysis and reporting over the last few years. The initial prototype was carried out in Excel and made use of the formulas for calculating a linear least squares fit and some related statistics. This tool remains an active part of CASA's tool set for demonstrating the methodology and illustrating outcomes to other CASA staff.

CASA then moved to storing data in an Access database and now has a specialist Data Warehouse. The quarterly reports are currently generated using Structured Query Language (SQL) code and this code has been provided to Data Analysis Australia for review. A summary of the data extraction and the regression analysis procedure as conducted by the excerpt of the SQL code is given below.

Data Extraction

A user defines the end date of the latest quarter of interest and the SQL generates a table of date ranges for the last twenty quarters counting back from the most recent. This table of date ranges is used for the extraction of relevant occurrences from the ASIR and ESIR data.

Two tables of safety incidents are created in the SQL: the ASIR Seed List, and the ESIR Seed List. The occurrences from the ASIR and ESIR table that fall between the date ranges given in the date range table are extracted. The contents of the two tables are as follows:

ASIR Seed List

- Occurrence ID;
- Quarter Number;
- Occurrence Type – a variable generated by concatenating the three description levels of an incident;
- Occurrence Location; and

- Case – this is the International Civil Aviation Organisation (ICAO) Type Designator (set to “ZZZZ” if the designator is undefined, and set to null if the occurrence ID is null).

ESIR Seed List

- ESIR Key;
- Quarter Number;
- Occurrence Type – taken directly from the ESIR data; and
- ESIR Operator – a variable generated by concatenating fields in two ESIR data tables.

Core Data

The two tables above are collated into a single table called Core Data, containing the following fields:

- Data Source – ASIR or ESIR;
- Quarter Number;
- Variable Type – State, Location, Aircraft, Occurrence Type and Operator;
- Variable Value – values within each variable type. For example, WA, VIC, QLD etc. in variable type “State”; and
- Occurrence Counts – Number of individual Occurrence IDs within each combination of data source, quarter number and variable value combination.

It can be seen from the above that the process used in the SQL to calculate the occurrence count is generating appropriate occurrence counts for each variable within each data source.

2.3.3 Regression Analysis

CASA carries out a univariate trend analysis for each variable reported on in the quarterly report. The SQL code was based on the analysis developed in Excel.

The analysis performed in SQL aggregates the incident counts by quarter and fits a linear model (note that the Excel tool looked at counts aggregated by month).

The SQL code takes the Core Data and performs calculations of the following for each combination of data source and variable:

- Regression intercept – with the occurrence counts as the response variable and the quarter number as the predictor variable;
- Regression slope – with the occurrence counts as the response variable and the quarter number as the predictor variable;
- R^2 – calculated by squaring the correlation between the occurrence counts and quarter number; and
- Standard deviation of raw occurrence counts, ignoring any linear relationship.

The results from the above calculations are formatted and collated into an Analysis Data table. Information presented in this table are as follows:

- Expected 1 standard deviation above – calculated as the sum of predicted occurrence counts and the standard deviation;
- Expected occurrence count;
- Expected 1 standard deviation below – calculated as the difference of predicted occurrence counts and the standard deviation;
- R^2;
- Standard deviation from expected – calculated as the ratio of residual (difference between the observed occurrence count and the predicted occurrence count) to the standard deviation;
- Gradient, i.e. Regression slope;
- Intercept, i.e. Regression intercept; and
- Standard deviation.

2.3.4 Interpretation

The trend monitoring component of the quarterly report flags deviations from expected trends. It identifies both higher and lower numbers of incidents than expected.

In the Excel version of the reporting tool there were four broad categories of analysis, exceedence indicators, trend statistics, lag one period correlation and a high variability flag. Now that the analysis is carried out using SQL only the exceedence indicators and one of the trend statistics (R^2) are calculated and reported.

Exceedence Indicators

The exceedence indicator is the number of standard deviations from the expected for the most recent observation. (Note that in Excel this was calculated for the most recent monthly observation and for the average of the three most recent monthly observations; now, using SQL, it is calculated only for the most recent quarter.)

The distance of an observed count from its expected value is expressed in terms of standard deviations. This is calculated as the observed count minus the expected count based on the fitted linear trend, then divided by the standard deviation of the time series. Note that the standard deviation was calculated based on the series itself, ignoring the order of the observations and hence ignoring any trend.

The exceedence values are colour coded for reporting purposes. The colours and thresholds used by CASA are given in Table 3.

This colour coding ensures that both higher than expected and lower than expected values attract reader attention.

Colour	Guide provided by CASA
Red	more than 2 standard deviations **above** the five year average/expected value – *a lot more occurrences than expected.*
Yellow	more than 1.28 standard deviations **above** the five year average/ expected value – *some more occurrences than expected.*
Blue	more than 1.28 standard deviations **below** the five year average/ expected value – *some less occurrences than expected.*
Purple	more than 2 standard deviations **below** the five year average/ expected value – *a lot less occurrences than expected.*

Table 3. Colour coding of exceedence indicators.

Other Statistics

In Excel, CASA implemented formulas to calculate four trend statistics, R^2, the F statistic, significance of the F statistic, and the annual change (as a percentage). It also implemented lag one period correlation and developed a high variability flag. Each of these is described below. Of all these, only the R^2 statistic is calculated in SQL.

The R^2 value is a measure of the proportion of the variation explained by the linear regression model, and is also equal to the square of the Pearson correlation coefficient between the observed counts and the times. In SQL the principle is the same but the implementation is on a quarterly basis instead of monthly.

The F statistic was calculated by the Excel procedure LINEST, which calculates the linear fit using the least squares criteria. The method implicitly assumes equally spaced measurements in time. The significance of the F statistic is calculated using the excel function FDIST. The arguments are the F statistic, one degree of freedom for the numerator and the length of the series minus two for degrees of freedom for the denominator.

The annual change was calculated in one of two ways, either using LOGEST in Excel or using a manual calculation implemented by CASA in the Excel workbook. The manual calculation is used only if the LOGEST calculation fails. LOGEST calculates an exponential curve that fits the data and returns the parameters. The parameters returned apply to a monthly change, so the figures must be annualised by raising the monthly growth rate to the power of 12, then subtracting one.

The manual calculation is given by

$$\left(ABS\left(\frac{E(S_n)}{E(S_1)}\right)\right)^{\frac{1}{n-1}}$$

where n is the length of the series, $E(S_n)$ is the expected value of the final incident count and $E(S_1)$ is the expected value of the first incident count. The formula gives the monthly annual change, which is then annualised.

The lag one period correlation is calculated using PEARSON in Excel on the series and the series lagged by one time period.

In the Excel workbook, CASA calculates a variability measure as the standard deviation of the series divided by the average of all the points in the series, also known as the coefficient of variation. If the variability measure is above 0.4 the high variability flag is set to Yes.

2.3.5 Investigation into flags

When a flag is raised, members of the Safety Performance Analysis Branch go back to the data and look at all the incidents of that particular type. The data contains fields such as the Occurrence ASIR Summary, not used in generating the analysis, and these additional fields can assist in explaining the observed trends and exceedences. If required, comment is sought from the appropriate field office before the information is included in the report for management.

2.4 The Outcome of the Report

A draft report is prepared and provided to management for consideration and eventual sign-off. From here the report is handed to two committees who may authorise further investigations. The report is used to highlight potential trends and not for any legal action, with the most probable outcome being an improvement in aviation safety. The tool is largely designed to keep track of what has been happening so that any unusual activity will not be neglected.

3. Comments on the Current Methodology

The current methodology is described in detail in Section 2.3. This section focusses on the exceedence indicators as these represent the core of the reported results.

The reported results display an observed count of the number of incidents in the most recent quarter and a corresponding expected count, presented for each operator, level three occurrence type, location, aircraft and state. The results are colour coded to highlight deviations from expected behaviour. This framework serves its main intended purpose, to highlight potential concerns in observed incident counts, and features several good points.

However, both the current methodology used to calculate the expected counts and the method for setting the thresholds leave CASA open to criticism for not following accepted statistical practice.

3.1 Good Aspects of the Current Methodology

- The Excel tool is sophisticated and makes full use of features like macros and named ranges.
- The analysis allows for linear trend.
- There is a process for identifying moderate and large deviations, both high and low, from the expected number of counts in the most recent time period.
- The attempt to allow for variability of the data in the calculation of exceedence values.

- Calculation of the R^2 statistic which indicates strength of the linear relationship.
- Use of an F statistic to test statistical significance of the linear trend in the Excel tool.
- The idea of reporting an annual percentage change.
- Use of a high variability flag in the Excel tool.
- Aggregation of the data (to monthly in the Excel tool and quarterly for the SQL code).

Data Analysis Australia has, for the most part, retained the underlying concepts in our recommended approach.

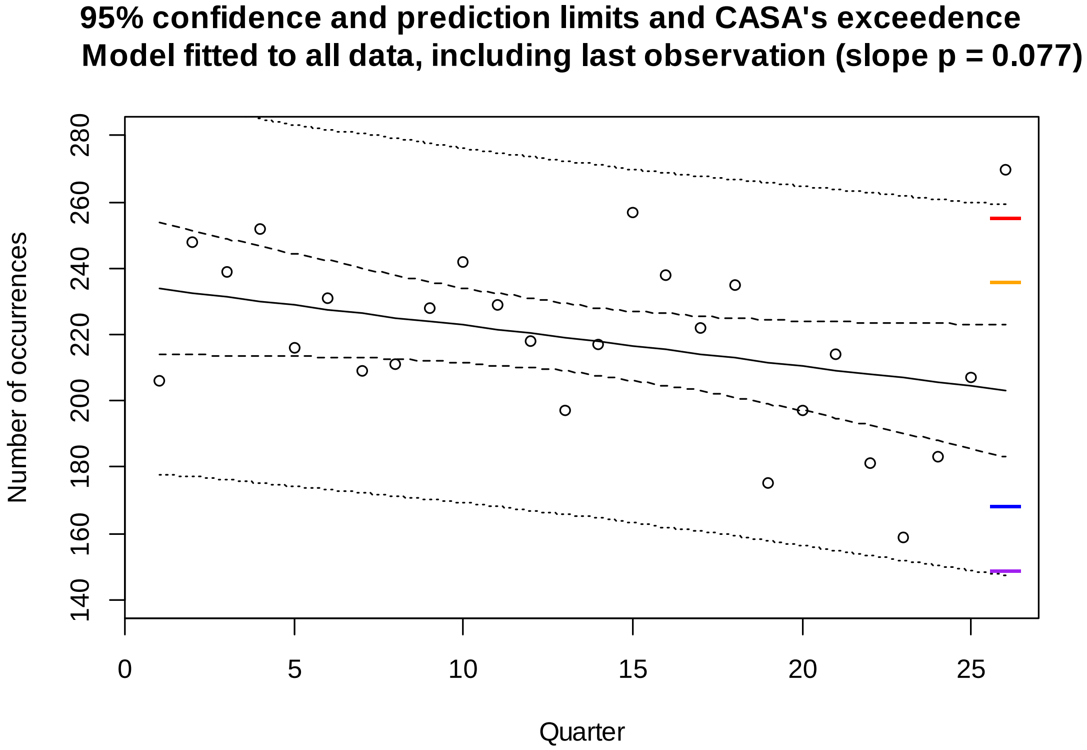
3.2 Deficiencies in the Current Methodology

- The linear regression implicitly assumes normally distributed data. Counts, especially small counts, are not normally distributed. A Poisson distribution is more realistic – it permits integers only, is never negative and allows for larger variability when the expected (average) count is larger. Section 4.6.2 includes illustrations of normal regression and Poisson regression models applied to two data sets. When the counts are large, it makes little difference, but with small counts there is a different outcome, and the Poisson regression is more appropriate.
- Ideally, when assessing whether a particular observation is atypical, the point in question would be removed from the dataset used to define the previous pattern. If the observation is extreme, its inclusion will influence the fitted line, particularly when it is the most recent observation in time series data. This principle is highlighted graphically in Figure 1. The last observation is much more strongly identified as unusual in the lower plot, where it has not contributed to the estimation of the fitted line.
- Using standard deviation calculated ignoring the linear trend is inappropriate. We want to compare the deviation of the observed count from the fitted line against the background variation *about the line*. This principle is highlighted graphically in Figure 2.
- The LOGEST calculation of annual change assumes exponential growth, i.e. a constant percentage increase each year, and as such is inconsistent with the linear trend analysis, although if the growth rate is small this makes little difference. If this calculation fails, for example because some quarter has no occurrences, an alternative calculation is performed, based on the fitted linear regression. However, if any fitted values are negative (which is unrealistic but can occur as a consequence of the model fitted, and further, is more likely to occur if there are no incident occurrences at the beginning or at the end of the period analysed), this method fails. An ad hoc solution is applied: ignore the negative sign and treat the value the same as a positive value of the same magnitude, for example treating a line going from –1 to 6 the same as a line going from +1 to 6. While this

enables a mathematical solution to be calculated, it does not give a sensible estimate of the growth rate.

Our recommended approach resolves many of these issues.

These points are elaborated in Section 3 in relation to several guiding statistical principles, with comments on how the current methodology performs relative to these principles. Appendix A summarises and compares some aspects of the modelling approach used by CASA and the Poisson generalised linear modelling approach recommended by Data Analysis Australia.



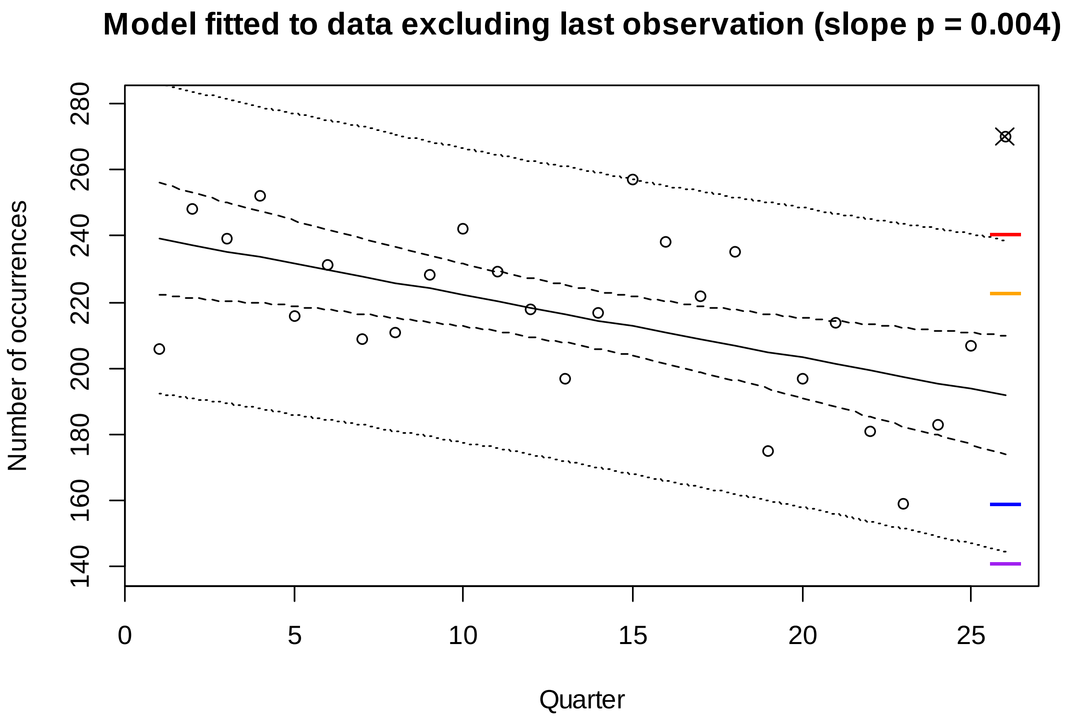


Figure 1. A demonstration highlighting the change in fitted line and exceedence indicators when the last observations is included (top) or excluded (bottom). The p-value quoted is for a statistical hypothesis test of whether the slope is zero. In the first plot, it is not significantly different from zero, at the 5% level, but is strongly significant in the second.

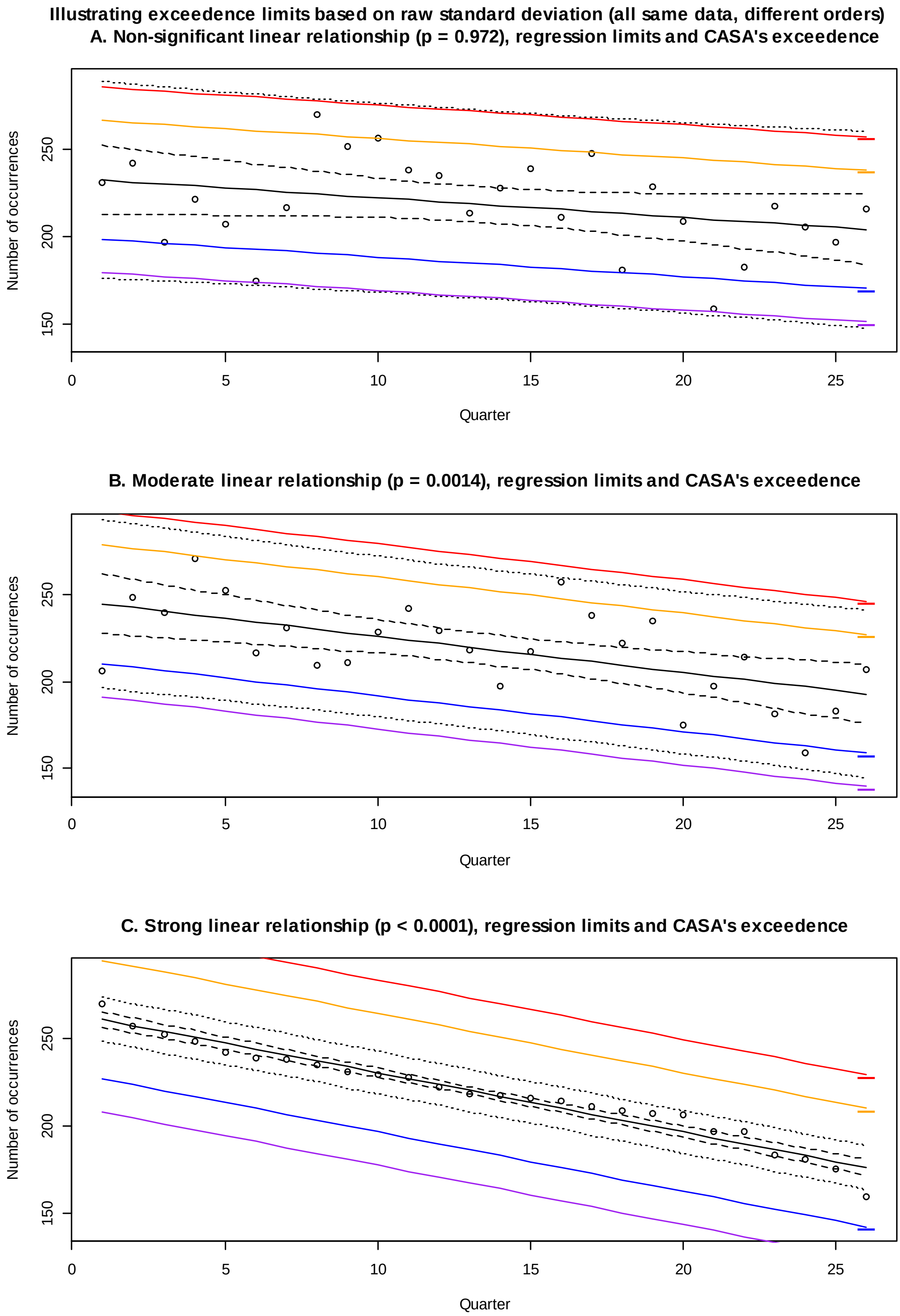


Figure 2. A demonstration highlighting the need to calculate the standard deviation based on the residuals rather than the data itself. The red, orange, blue and purple lines and marker for the last observation relate to the threshold colours in Table 3, while the dashed and dotted black lines are 95% confidence bands for the fitted line and prediction interval for the individual point respectively. The dotted lines better indicate potentially unusual observations.

3.3 Some Guiding Statistical Principles

3.3.1 Generalised Linear Modelling for Count Data

Safety incident data is count data. Therefore, it would be appropriate to assume that the data comes from a discrete distribution. The Poisson distribution is often used to model count data. The Poisson distribution gives the probability of k events occurring in a specified time interval (and assumes that the times between events are independent). It is appropriate even if the expected number of events is small.

The consequence of assuming that aviation safety incidents have a Poisson distribution is that generalised linear models must be used in the place of ordinary linear regression. The generalised linear modelling framework allows for the response variable to have a distribution that is not normal. Each distribution has a natural link function (a transformation applied to the expected value of the response variable to ensure fitting a linear model is reasonable) and the natural link function for the Poisson distribution is the logarithmic link. A linear model is fitted to the logarithm of the expected count at a given time, with the observed counts assuming to follow a Poisson distribution whose mean is equal to that expected mean.

The benefit of explicitly allowing for the response to have a Poisson distribution is that there are standard tools and procedures that can be used for analysis, including those implemented in the freely available software package R.

The current method for calculating the expected counts assumes that the data is normally distributed and it also calculates exceedences. These are discussed in turn.

In general, for events for which there will frequently be zero occurrences, the counts do not follow a normal distribution. Therefore, as described above, analysis of count data typically assumes that the data comes from a Poisson distribution and Data Analysis Australia recommends that the Poisson distribution is assumed in the next iteration of the methodology for analysis of safety incidents.

The current method for evaluating if an observed point is of concern takes the observed count minus the expected value and divides it by the standard deviation calculated on the entire series, uncorrected for any trend. Using the standard deviation of the series itself is not considered good statistical practice. Generally the variability measure used is based on the residuals (that is the observed minus the expected values), taking into account the linear trend if present.

Therefore, calculating the expected counts based on a linear trend with data assumed to be normally distributed (particularly with the current method of determining the threshold) may highlight some behaviours which are not of concern and may miss some behaviours that are of concern.

3.3.2 Identifying Different Sources of Variation

Many different aspects can contribute to variation in observed data. A common aim is to identify and allow for different sources, and to flag for further attention sources of variation that might be of concern or interest. Such explanations for variation could be linear trend, the last observation is different or special (that is, it departs from the earlier pattern), clusters, seasonality, or different exposure to risk of the incident.

3.3.3 Obtaining Consistent Results

The suite of analyses used should result in a consistent view of concern about aviation safety. This is best achieved by using one unified modelling approach and reporting on trends and statistics that result from the modelling process.

Currently there are two methodologies for calculating the annual change; this may lead to inconsistent views of concern.

3.3.4 Recording and Detecting Clusters

There are two forms of clusters that may occur in aviation safety incidents.

Firstly, one event might generate more than one incident (for example, if there was insufficient separation between two aircraft then there at least two incidents would be reported). This is correctly accounted for during the data setup, where multiple records may be associated with a single occurrence id.

Secondly, behaviours like systematic failings in maintenance might lead to a cluster of incidents. This should be detected in the trend analysis as proposed by Data Analysis Australia as an unusual spike in the relevant category; however, it does highlight the need to perform analyses both on individual occurrence types and on aggregates by meaningful categories such as geographical locations.

3.3.5 Generalised Linear Modelling with an Offset

A key consideration of the trend analysis is understanding the expected behaviour. The assumption of a linear trend over time may be appropriate in some circumstances. If, for example, there was a linear increase in the number of flights departing from a particular location then it would potentially be acceptable to see a corresponding linear increase in incidents at that location over time. However, if the number of flights departing remained constant but the number of incidents increased in a linear fashion this may be considered cause for concern. Therefore, the trend analysis should compare the observed trend relative to the expected trend.

The generalised linear model assumes a linear trend (albeit on a logarithmic-transformed scale). One key consideration in the analysis of safety incidents is not just the incident itself, but the number of incidents relative to opportunity. Opportunity for incidents, or exposure, might be difficult to measure. If it was possible to develop an exposure measure the generalised linear model framework outlined above can be extended to include this as an offset. The offset is a measure of exposure per observation (in this case, per quarter). Then what might seem like an unchanged trend (and therefore of no concern) in a particular incident class might become a concern if the potential for that type of incident is declining. Alternatively, an apparent increase in the number of incidents might be simply explained by increasing exposure – the number of occurrences per unit exposure might not have changed. However, if, after allowing for exposure, there is still an increasing trend in the number of incidents, this might give additional cause for concern.

A mathematical description of the inclusion of an offset for Exposure is provided in Appendix A.

4. Specific Recommendations

4.1 An Alternative Methodology

Data Analysis Australia agrees that the primary focus of the quarterly reporting should focus on concerns about recent events. Therefore, the recommended methodology describes a way to highlight when recent observations are different from past experience.

The methodology recommended by Data Analysis Australia:

- Uses a generalised linear modelling framework;
- Assumes the count data comes from a Poisson rather than a normal distribution as this is much more realistic, especially when the counts are small;
- Uses a logarithmic link, modelling the logarithm of the expected counts as a linear function of time and thereby precluding negative expected counts;
- Evaluates whether the most recent observation is consistent with expected behaviour based on previous observations;
- Examines other patterns or sources of variation in the data, such as non-linear trend, changes in level, and seasonality (and flags these as appropriate);
- Detects and flags overdispersion, and modifies the analysis to allow for it;
- Detects and flags underdispersion; and
- Can also include an exposure measure, if available, to account for additional variation and estimate the rate of incidents per unit of exposure. (If the exposure is changing linearly over time, the time variable will act as a proxy for exposure, but an independent measure of exposure is much preferred.)

4.1.1 Generalised Linear Modelling Framework

When data arises from counts of random rare events it typically has a Poisson distribution and is therefore usually modelled using a generalised linear modelling framework, with a logarithmic link function. Data Analysis Australia recommends moving from a linear modelling framework to a generalised linear modelling framework. The approach can be called Poisson regression.

The primary model Data Analysis Australia recommends fitting has two terms, a linear trend term and a `final observation' effect. The final observation effect is an indicator variable which is one for the final observation and zero elsewhere and will be significant if the final observation is `special' in some way and doesn't fit in with the pattern established over the proceeding time periods. Once this model is fitted, model-selection techniques are used to determine the final (best-fitting) model among special cases of this model – the simplest model that adequately describes the data. The final model would be one of four possibilities:

M1. A model with both a linear trend and final observation effect;

M2. A model with a final observation effect;

M3. A model with a linear trend; or

M4. A constant model.

The relationships between these models are displayed in Figure 3. An arrow connecting a model A to a model B indicates that model B is a special case of model A – we say model B is nested within model A, and a likelihood ratio test can be performed to compare the models, or, equivalently, to test whether the term that has been dropped is statistically significant.

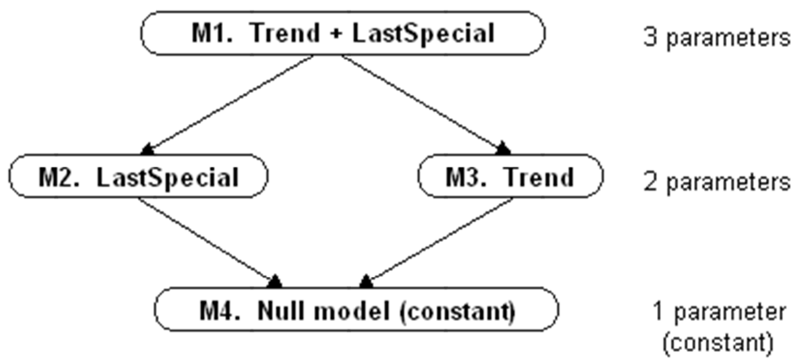


Figure 3. Lattice diagram of nested models for model selection. An arrow from a model leads to a simpler model that is a special case of the first model. Likelihood ratio tests can be used to compare these two models.

Starting with the most general model, M1, paths are followed through the lattice, applying likelihood ratio tests until a statistically significant difference is encountered, implying the model cannot be further simplified. The best-fitting model is the simplest model that can be reached in this way.

If either M1 or M2 is found to be the best-fitting model (we cannot proceed to M3 or M4), this means that the LastSpecial term is statistically significant. This suggests that the final observation is inconsistent with the pattern established by the prior observations, and therefore warrants attention. The level of statistical significance would be used to highlight the level of concern (and both higher and lower counts than expected can be identified). P-values for a given effect are obtained from the likelihood ratio test between models including and excluding the relevant term.

4.1.2 Understanding High Variability

The Excel tool built by CASA has a High Variability flag. The flag will typically highlight if the series has one of the three sources of variability, namely:

1. The series has a higher level of variation than expected;

2. The series has a strong trend; and
3. The series has a lot of zeros.

Data Analysis Australia proposes that each of these be identified separately.

One of the theoretical properties of the Poisson distribution is that the variance is equal to the mean, though this is not always the case in practice. If the variance is similar to the mean the Poisson distribution is appropriate; however, if the variance is statistically significantly greater than the mean, the data is considered overdispersed (to have variation over and above the expected level of variation for Poisson counts).

For such data, the Poisson analysis can be modified to a quasi-likelihood approach, in which the surplus variation is allowed for and F-tests replace the usual likelihood ratio chi-square tests.

Conversely, if the variance is statistically significantly smaller than the mean, the data is considered to be underdispersed. Typically the Poisson distribution is used but these situations may be worth flagging as the data is more consistent than expected, which might raise suspicions about the incident-reporting process.

Data Analysis Australia also recommends that statistically significant linear trends be highlighted. This would be particularly relevant when an offset for exposure is not used so that CASA is aware of which series are experiencing increases, decreases and no changes over time.

Data Analysis Australia also recommends highlighting when a large proportion of the quarterly observations are zero. A statistical test can be performed for whether there is "zero-inflation", i.e. surplus zeros beyond the number expected from a Poisson distribution.

The suite of models fitted and compared is extended in Appendix B to allow further investigation of trends and sources of variation.

Occasional unusually large or small counts are called outliers, and can contribute to large variation and reduce the power of statistical tests to detect patterns. An illustration is given in Section 4.6.2 (Figure 7). It is recommended that the presence of gross outliers be noted as a warning.

4.1.3 Modelling Additional Sources of Variation

Various extensions of the models described in Section 4.1.1 can be investigated to see if alternative explanations of the data are feasible. Some of these might raise flags for further investigation, or they might alleviate concern by providing a feasible alternative explanation for the patterns observed. Examples are: nonlinear trend, a change in level before the last time period, and seasonality.

In some cases overdispersion that is apparent in model M1 might disappear with a different, more general and more appropriate model. In other words, the overdispersion has been explained by including different terms in the model and the overdispersion flag identified in the primary analysis can be disregarded. However, for consistency, we suggest that the overdispersion flag already raised be left unchanged, but the behaviour noted.

4.1.4 Adjusting for Exposure

A generalised linear modelling framework can include an offset. In this context the offset could be a measure of the exposure to the possibility of incidents (for example there could be a decreasing trend in exposure due to an aircraft type being phased out of commercial usage). Including exposure as an offset in the general linear model effectively allows a rate of incidents per unit exposure to be modelled. Appendix A gives more details. Allowing for exposure in this way provides greater sensitivity for detecting patterns or trends of real concern or interest.

Data Analysis Australia recommends that the issues of exposure are considered, particularly to determine if any of them are measurable.

4.2 Consistent Analysis and Reporting

Data Analysis proposes that primary flags be based on the models M1 to M4. These primary flags are for:

- The last observation appears special, and whether large or small;
- A statistically significant linear trend, and whether up or down; and
- Overdispersion or underdispersion.

With the addition of different terms and different models, some of these flags might change. In these circumstances we propose to note this but not change the primary flags. If this proves to flag too many cases to follow up individually, this approach could be reconsidered.

4.3 Aggregation of the Data

When counts are very small, even though the Poisson distribution can be expected to be much more appropriate than assuming a normal distribution, it can be very difficult to detect if an occasional larger value is out of the ordinary, or within the typical distribution of values expected for Poisson data with a very small mean.

The current methodology aggregates the data by quarter. Data Analysis Australia believes that this choice represents a good trade-off between having low counts and detecting concerning trends without a large delay. It also ensures the data can be essentially complete before analysis.

However, CASA should be mindful of the possibility of this aggregation obscuring some cycles, for example day of week. If a particular incident is more likely for non-commercial flights they may also be more likely on the weekends. A quarterly (or even monthly) based aggregation might obscure evidence of a weekly cycle.

4.4 Number of Quarters to be Analysed

Data Analysis Australia recommends that a whole number of years be analysed, and suggests either five years (20 quarters) or six years (24 quarters). The suggested models fit up to three parameters, or up to five with seasonality included, and $n = 20$ is a suggested minimum for this; $n = 24$ would be preferable, but must be tempered with validity of the data, consistency of definitions and reliability of reporting, and the likelihood of consistency of patterns or trends over that period. However four years ($n = 16$) or fewer observations is unlikely to give sufficient power to detect changes in patterns.

4.5 Further Approaches, Including Multivariate

There are many different approaches and techniques that could be used. Our aim is to keep it relatively simple while still giving a good chance of detecting worrying trends or patterns, without too many "false positives", i.e. flags raised for further attention that turn out to be of no concern.

The approach recommended considers each occurrence type in isolation. The same approach can be used for occurrence types aggregated by operator or geographical region, for example. It is expected that users of the outcomes would notice consistent patterns, for example two or more occurrence types that increase together. They could also be expected to be aware of the types of occurrence types that might increase in parallel, perhaps with a common cause. Some multivariate approaches or simple correlation analysis might be possible, comparing outcomes for several occurrence types simultaneously, but these have not been considered in this report.

4.6 Reporting Outcomes

The analyses recommended in this report should be regarded as a screening tool with various aspects of the data flagged or highlighted for each occurrence type, or within different geographical regions. The outcomes can be compactly summarised in a table, with one line for each occurrence type, with further descriptive information in graphical plots which would be referred to only for those occurrence types seen to be of interest from the table.

4.6.1 Table of Outcome Flags

Data Analysis Australia suggests that the outcomes of the analyses be reported in a table such as shown in Appendix C (Figure 13), with flags indicating different aspects that are of interest or potential concern, requiring further investigation. Flags can indicate changes in pattern that might suggest possible safety issues or improvements in safety, or, together with further analysis, might indicate other characteristics that can be satisfactorily explained in other ways and raise no concern or special interest.

Suggested flags are defined with more detail in Appendix C, but brief names or descriptions are listed here:

- LastSpecial, and whether high or low;
- Trend, and whether increasing or decreasing;
- Second-last observation special;

- Pattern changes when last observation added;
- Underdispersion;
- Overdispersion;
- Excess zeros;
- Outliers;
- Change in level;
- Nonlinearity;
- Seasonality – annual cycle;
- Dispersion changed by other terms;
- Trend in rate of incidents per unit exposure.

Two quantities calculated in START but with no analogue in the recommended approach so far described are annual percentage change and a lag one autocorrelation coefficient. CASA had advised that they do not wish to consider these at present, but to bear them in mind for the future. If used, these could be included as two additional columns between Overdispersion and Excess zeros:

- Annual percentage change (based on linear model, M3);
  - Probably only report this if there is a significant linear trend;
  - Can be directly calculated from the Poisson model with logarithmic link, more naturally than from the linear model fitted by START, and without the dilemma and inconsistent calculations caused by zero counts;
- Lag one autocorrelation coefficient;
  - Not ideal for Poisson data, especially with small counts;
  - Based on residuals from the linear trend model, M3;
  - Suggest using square root of the counts minus square root of the fitted values;
  - Note, if autocorrelation is present, the assumption of independence from observation to observation is false. This reduces the validity of other statistical tests performed, although they can still be a useful screening device.

Not all of these flags need be retained. CASA is invited to suggest those they think might be dispensed with, giving reasons, or to suggest others.

For example, there are two characteristics reported in the START spreadsheet that have not been included here. They are the lag one correlation, to check for serial correlation among counts of incidents, and a percentage annual change in the number of incidents. Analogous indicators could be calculated here if thought to be valuable. In fact, since the Poisson model is fitted with a logarithmic link, the annual percentage change is a much more natural and direct statistic to derive than from the normal linear model fitted in START. For the serial correlation, it is suggested that this be calculated on the square root of the counts, which will approximately stabilise the variance.

The aim is to highlight aspects that might be suggestive of safety concerns, but not to highlight too many that turn out to be non-issues. The approach could be tested first on historic data and then on an ongoing basis to help calibrate the best level at which to set the flags. It is preferable to begin with more flags and drop out those that appear not to be useful than to leave out too many in the first instance.

For ease of inspection we suggest symbols to aid quick interpretation, such as displayed in the example table in Appendix C.

Remember that this is a *screening* tool. We have aimed to include a large number of flags that *might*, in some cases, be associated with changes in the pattern of safety incidents so that they can be investigated further. We hope that we have included enough characteristics to detect any issues of concern, but there is no expectation that every flag raised will relate to a safety issue. Identifying the issues that are of concern, from among those flagged, is the purpose of the next stage of the CASA's process.

The aim here is to capture all those situations that need followup while not capturing too many "false positives". There will always be a tradeoff. It is hoped that the breakdown of the variation in the data into various aspects will assist in quickly determining which occurrence types need followup. If an unacceptably large number of false positives arise, adjustments can be made to the procedure.

4.6.2 Plots to Aid Interpretation

Another excellent aid to interpretation is a plot of the data together with fitted models represented by lines. Although many fitted models are considered in the course of the analysis, only a few will be relevant. It makes sense to tailor the plot for each incident type analysed to show only a few most relevant models.

Two procedures are possible:

- A plot with the models deemed most interesting or relevant for each variable is produced for every occurrence type and stored in an electronic file but only accessed if required, as determined by CASA staff members doing the checking. It would be particularly nice to be able to have a link from the table of outcomes to display individual plots on request – but the feasibility of this is uncertain. This would be the simplest, but means the particular models displayed would be determined in advance rather than be able to be chosen by the user in any particular instance. However it should not be too hard to choose a few models (perhaps up to four) that include all the models likely to be of interest in any given case.

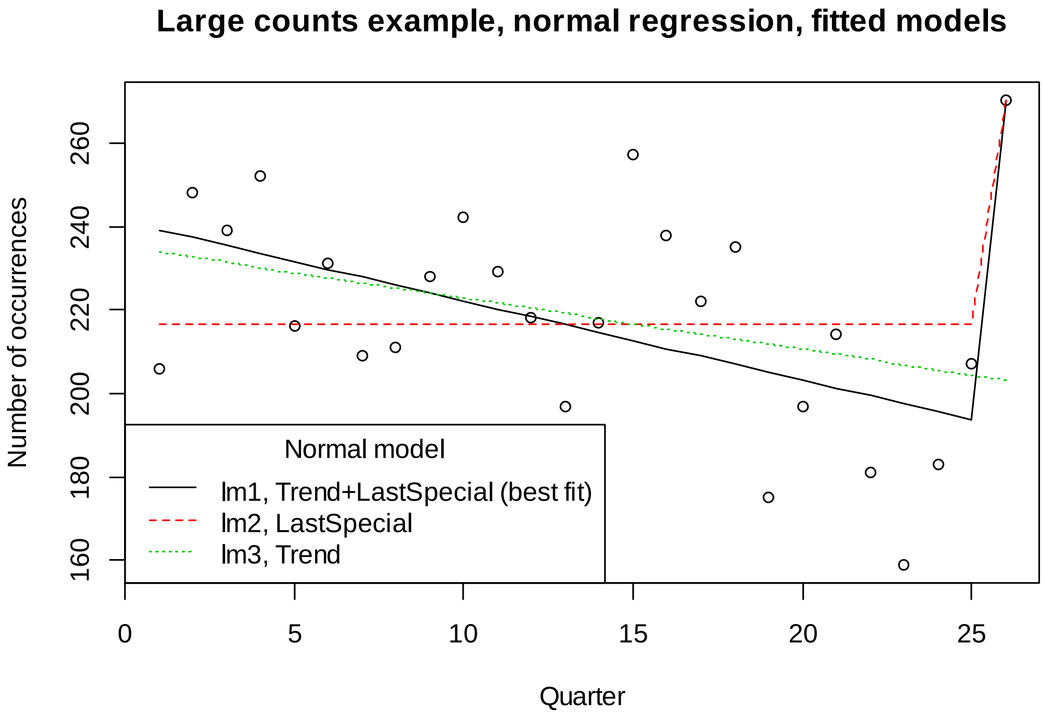
- An interactive process whereby the user can say what terms they wish to include in the model displayed. For example, the user might wish to view a model that includes just a change in level, another model that includes linear trend, and another that includes linear trend and seasonality. This would require considerable time and effort to make it intuitive and seamless. Perhaps a better alternative if this capability is really required would be for a CASA staff member to learn how to fit the models in the software package R[1]. Some basic instructions and examples could be provided.

Figure 4 and Figure 5 show some examples of plots of the data with three fitted models (M1, M2 and M3) overlaid – for both the old normal regression model (top plot) and the Poisson regression (bottom plot). There is little difference between the analyses in the large counts example (Figure 4) but the two approaches yield different conclusions in the small counts example (Figure 5).

Figure 6 for Overweight landing (data slightly changed for illustrative purposes) shows an example with a change in level model (M5 in Appendix B) and a quadratic trend model (M6 in Appendix B). For this variable, the best fitting model among M1 to M4 was M3, linear trend. However both the change in level model (M5) and the quadratic trend model (M6) fit better. Inspection of the data in Figure 6 helps make sense of this.

Models M5 and M6 cannot be directly compared because they are not nested, but because they have the same number of parameters, we can compare them on the basis of their residual deviance. Deviance is the generalised linear modelling analogue of residual sum of squares, and smaller deviance indicates a closer fit to the observed data. The change in level model has smaller deviance so is said to have the better fit. However the quadratic model is also a contender, so it is worth reporting both and flagging both change in level and possible nonlinearity. It will require further (future) data to determine whether there really has been a downturn (as per the quadratic model), or whether the changed level is sustained.

[1] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.



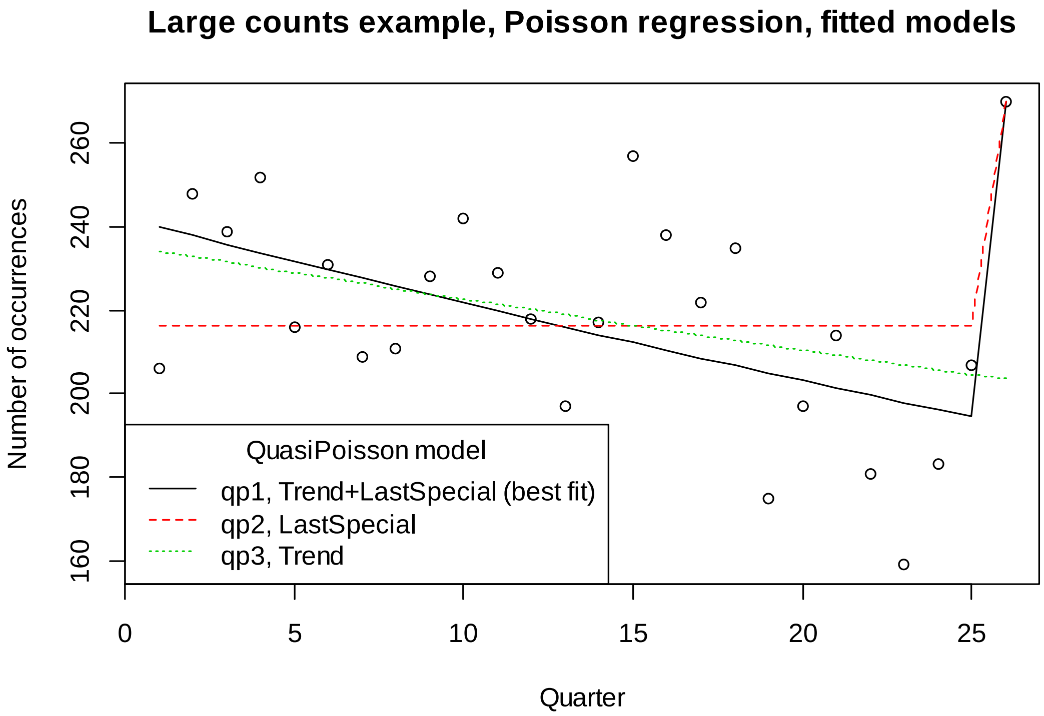
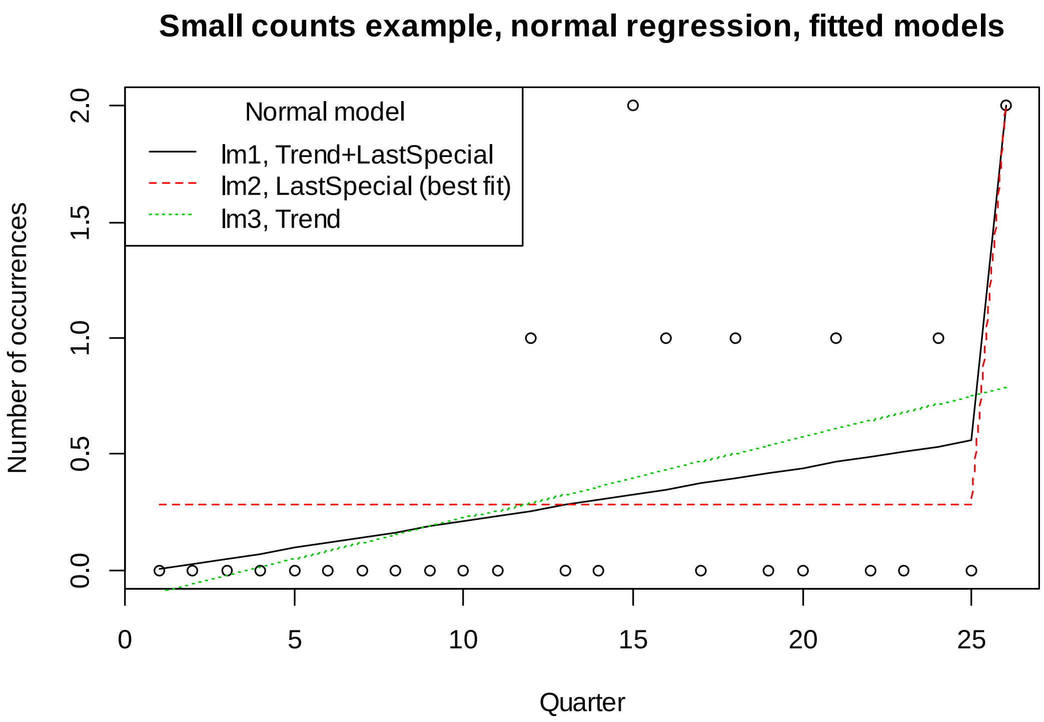


Figure 4. A demonstration based on a large counts example. Both the normal regression model and the Poisson regression model highlight a downwards linear trend and statistically significant increase in incident reports in the last quarter.



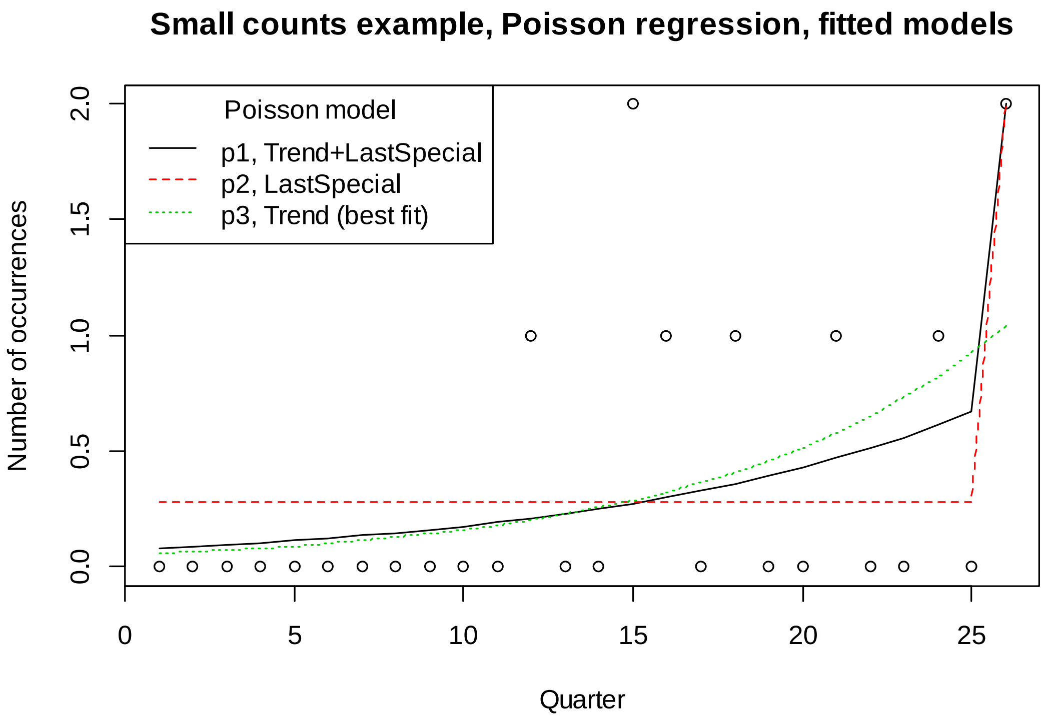


Figure 5. A demonstration based on a small counts example. The normal model best fit suggests no linear trend prior to the last observation, but a different expected count in the last quarter. The Poisson model best fit suggests that a linear trend over all 26 observations adequately describes the data – in other words, the last observation is *not* identified as unusual relative to the pattern established earlier.

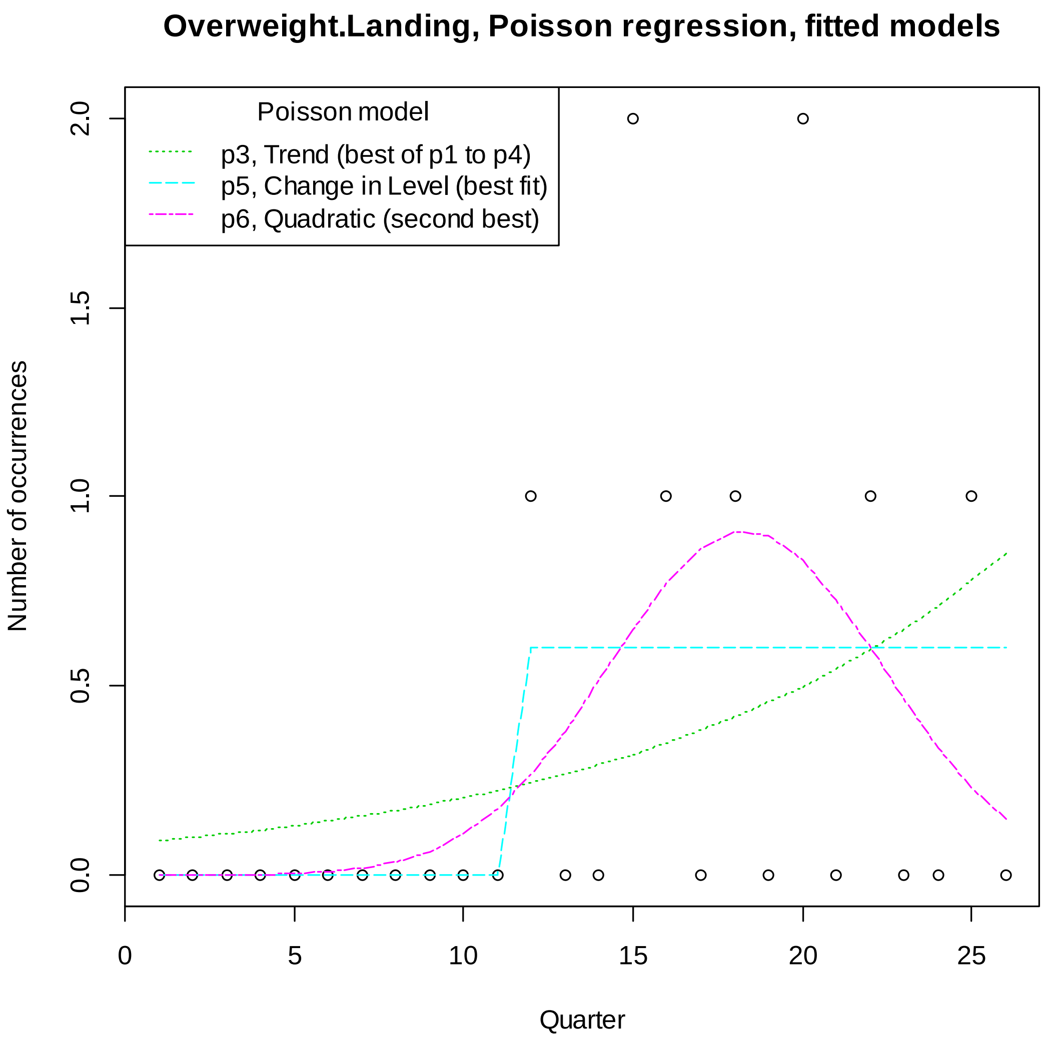


Figure 6. Data and fitted models for Poisson models p3 (linear Trend), p5 (Change in level) and p6 (Quadratic trend) for Overweight.Landing.

The Runway Incursion data in Figure 7 exhibits overdispersion and therefore the analysis was modified by using quasi-Poisson models. However the principles remain the same. In this example, the best fitting model among M1 to M4 was M1, constant; linear trend was not statistically significant. However there is a gross outlier evident in Figure 7. Further investigation revealed that if the outlier was removed from the analysis, the linear trend (model M3) became significant. This illustrates how the presence of an outlier can adversely affect the analysis – the variation added by the outlier reduces the power of the statistical tests to detect other patterns.

It is not feasible to pay such individual attention to each variable in an automated procedure as is required. However we can flag the presence of outliers, which indicates that other tests might be less reliable. For this reason, a flag for gross outliers is suggested.

Both the change in level model (M5) and the quadratic trend model (M6) fit better than the constant model (M4). Inspection of the data in Figure 7 suggests that the outlier, as well as weakening the statistical test for linear trend by adding variability, also strongly influences the quadratic trend. If the presence of an outlier is flagged, it is recommended that the plot be inspected, even if no other flags are triggered.

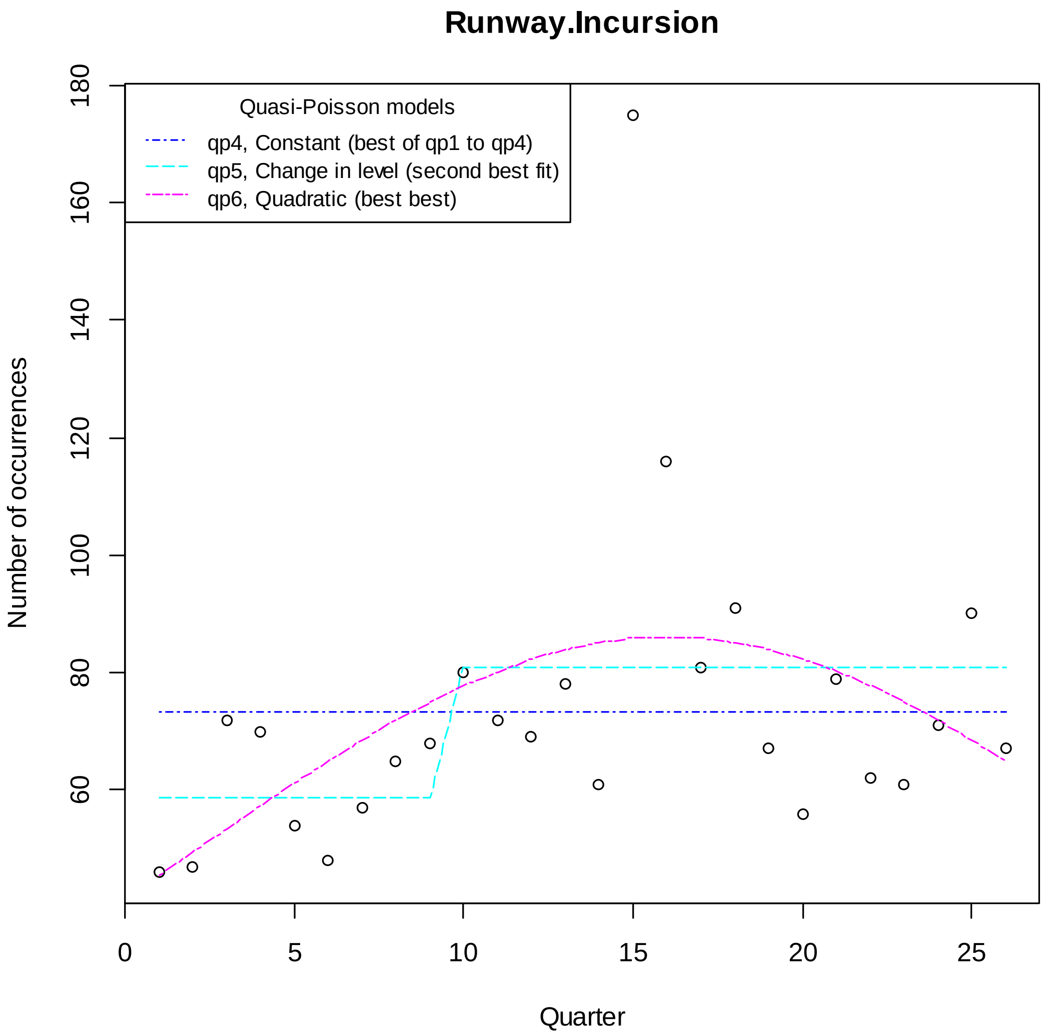


Figure 7. Data and fitted models for quasi-Poisson models qp4 (constant), qp5 (Change in level) and qp6 (Quadratic trend) for Runway.Incursion.

Figure 8 shows another example for which the quasi-Poisson model was required. In this example, the best-fitting model was the change in level model. There was also evidence of excess zeros, based on a test with a zero-inflated Poisson model. It is possible that there was under-reporting in quarters 6 to 9.

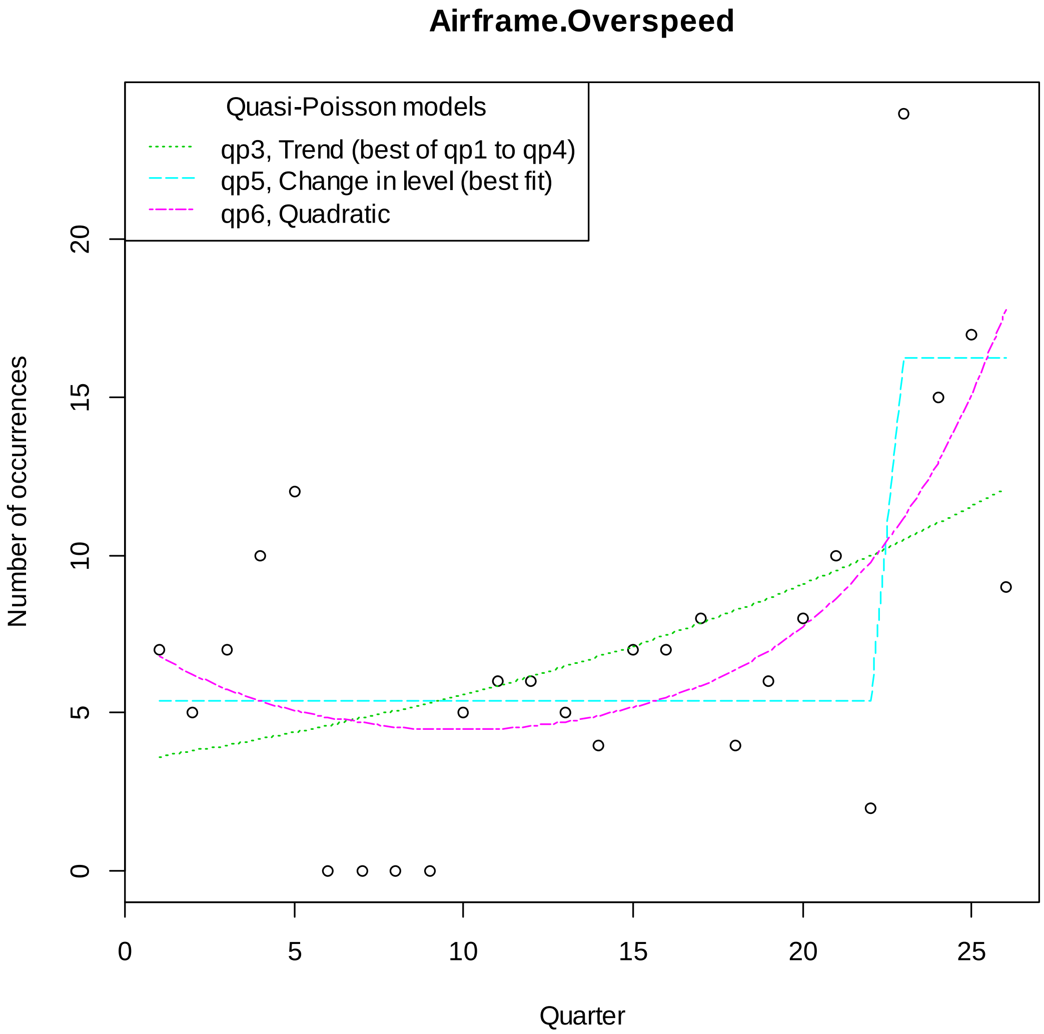


Figure 8. Data and fitted models for Aircraft.Overspeed.

Seasonality was not found to be significant in any of the examples we tried, but we present an example in Figure 9 to illustrate the shape of fitted models. The top plot shows a simple annual seasonal model, with the same cycle repeated each year. The second plot shows a linear trend plus a seasonal cycle. The oscillations get larger as the mean increases. This is a consequence of the logarithmic link. On the logarithmic scale on which the linear predictor is estimated, the oscillations have the same height, as can be seen in the bottom plot of Figure 9. Similarly it can be seen that the linear Trend truly is linear here, but a slight curve is introduced when it is back-transformed to obtain the fitted values for the linear Trend model.

A further example was contrived to exhibit both trend and seasonality. The data and some fitted models are shown in Figure 10.

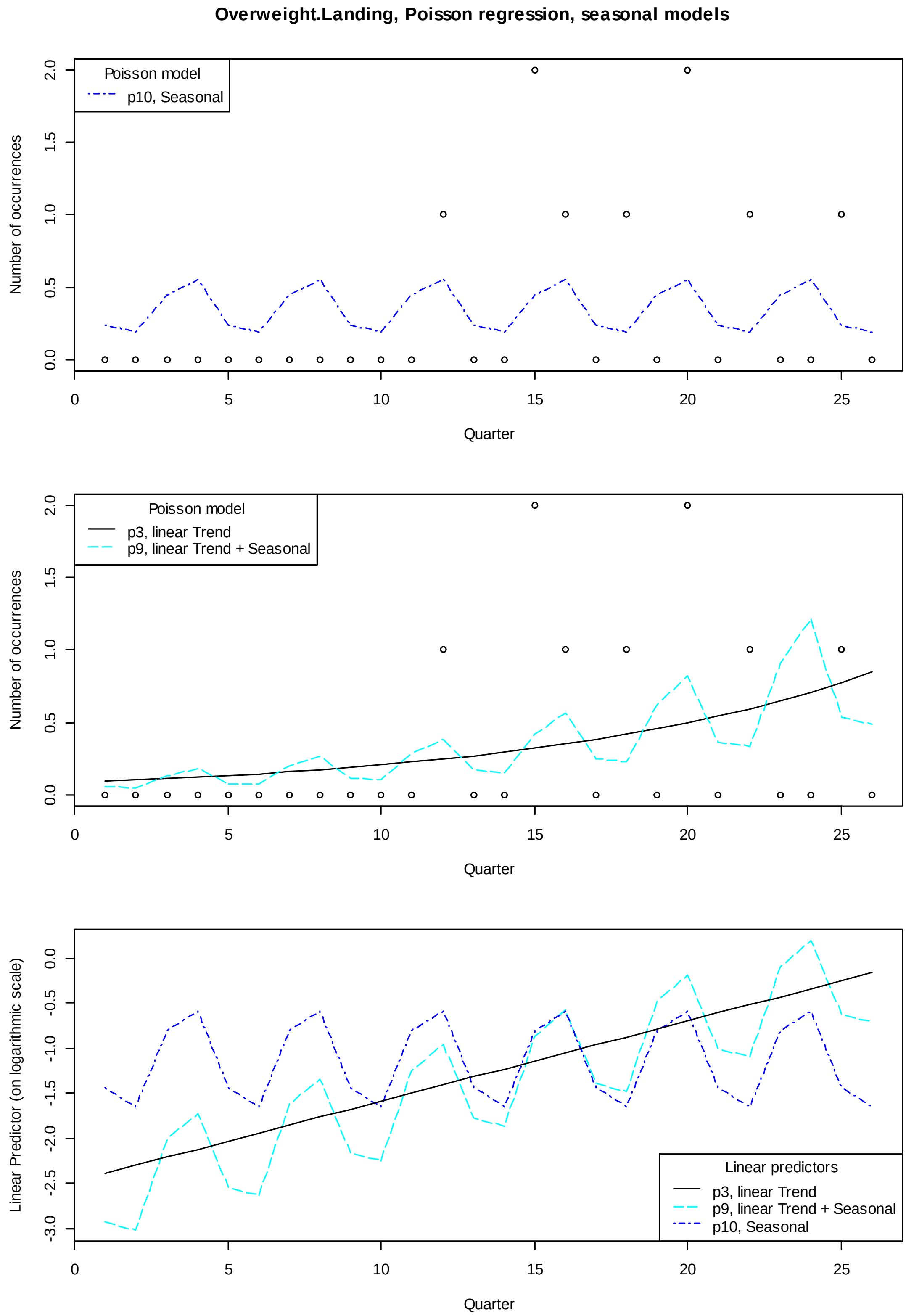
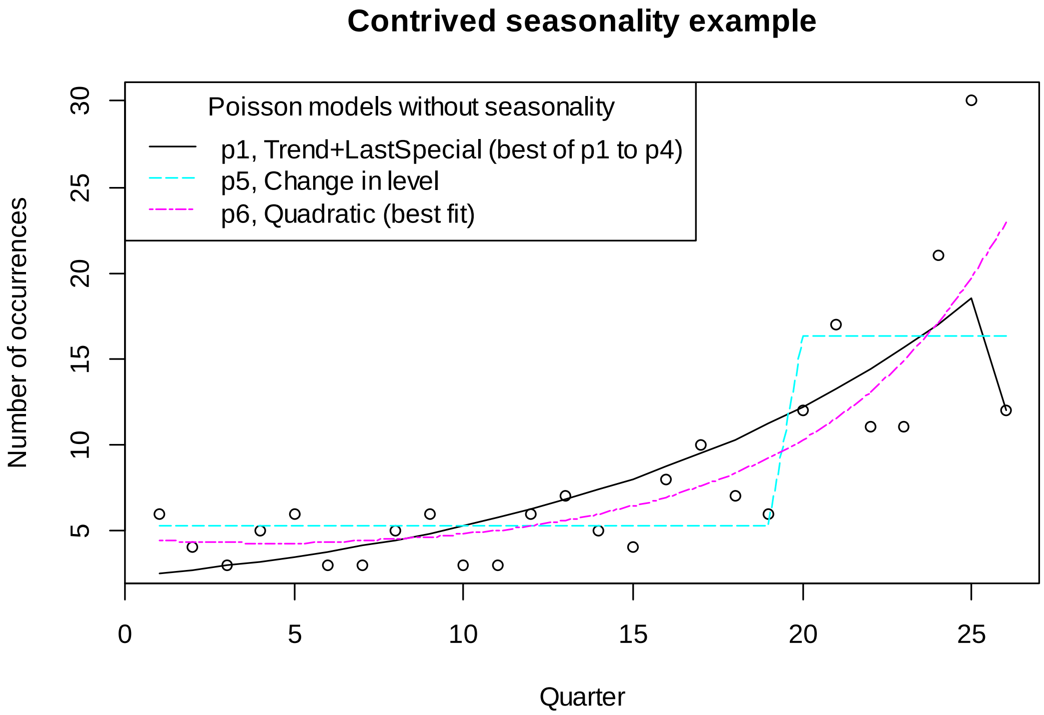


Figure 9. Examples of some seasonal models (not statistically significant here), and the linear predictor on the logarithmic scale. The fitted values in the top two plots are the exponential of the curves displayed in the bottom plot.



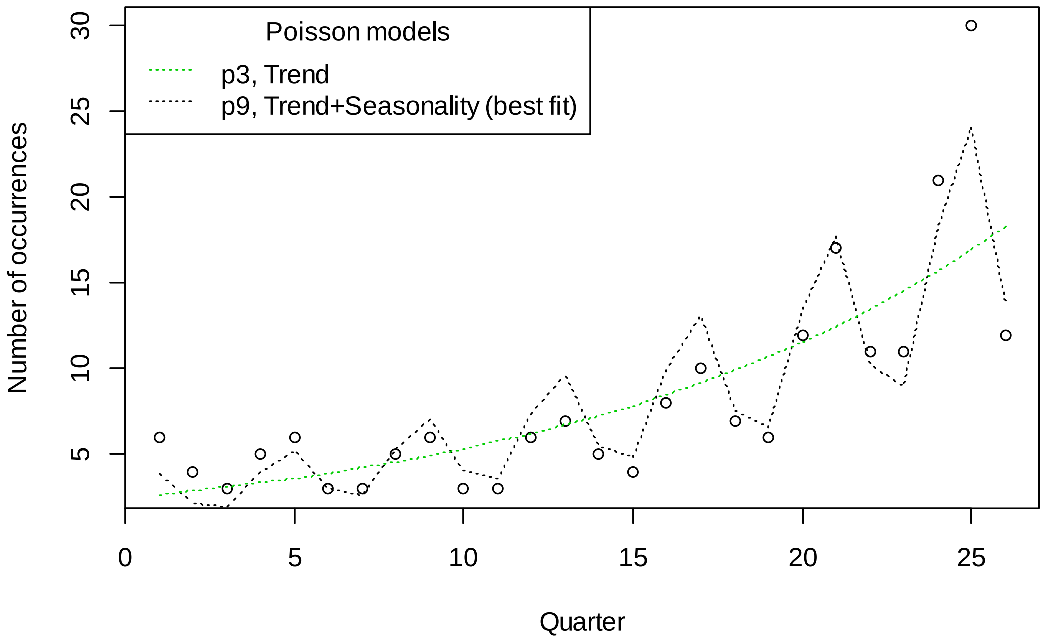


Figure 10. Contrived example exhibiting both trend and seasonality.

5. Summary

Data Analysis Australia recognises several good characteristics in the safety analysis procedure currently used by CASA, and seeks to preserve these ideas while simultaneously improving the analysis by fitting generalised linear models based on a Poisson distribution for the observed counts of incidents.

Data Analysis Australia recommends a revised safety analysis procedure in which, for each occurrence type, a number of simple models are fitted to quarterly counts over the previous five or six years and compared to identify patterns and trends. Different types of changes are identified and flagged in a summary table. This is a screening tool to highlight occurrence types that warrant closer investigation because they potentially indicate safety problems, or improvements, or other changes of interest. Plots of the data and selected fitted models can greatly aid interpretation.

Only univariate analyses have been considered – one occurrence type at a time. The same techniques can be applied to data aggregated over multiple occurrence types, within one operator or aircraft type, for example. It is expected that users of the outcomes would note consistent patterns emerging from multiple occurrence types that might stem from a common cause.

Appendix A. Summary of Modelling Approaches

The CASA Linear Regression Approach

Linear Model

$$Y = \alpha + \beta t + \varepsilon, \ \ \varepsilon \sim N(0, \sigma^2_{Y.t})$$

where ~ can be read as "is distributed as" and $N(\text{mean}, \text{variance})$ denotes the normal distribution with the stated mean and variance. This model can alternatively be written as

$$Y \sim N(\alpha + \beta t, \sigma^2_{Y.t})$$

Y is the observed count at time (month or quarter) t, α is the intercept and β is the slope, parameters which are estimated from the data with the estimates denoted a and b respectively. The variance about the regression line $\sigma^2_{Y.t}$ can also be estimated. This estimate is different from the square of the naïve standard deviation calculated in the CASA approach,

$$\text{standard deviation} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(Y_t - \bar{Y})^2}\ .$$

CASA's standard deviation ignores any trend or pattern in the data.

The expected count for any particular time (month or quarter) t is the point on the line, $\alpha + \beta t$, estimated by $a + bt$. There is nothing in this model to prevent expected counts from being negative.

The Poisson Generalised Linear Model Approach

$$Y \sim \text{Po}(\mu(t))$$

where Po(mean) denotes the Poisson distribution with the given mean. We model the mean as a function of time t with a logarithmic link function, which ensures expected counts are never negative.

$$\ln\big(\mu(t)\big) = \alpha + \beta t$$

where again we estimate the intercept and slope by a and b. Under the Poisson model, the variance of the observed count Y at any given time t is equal to its variance at that time. Sometimes the variance in the data is larger than this. This is called overdispersion. Occasionally the variance is smaller than expected, called underdispersion.

Dispersion can be statistically tested, and if overdispersion is found to be present (often reflecting additional unexplained variation), a variant of the analysis called a quasi-Poisson model can be fitted. Underdispersion should raise a flag indicating attention is required. It suggests less variability than would be expected for random counts and may suggest improper reporting.

We can also test the strength of the linear relationship (whether β could be zero), and whether non-linearity is present.

Further, we can test whether the most recent observation is inconsistent with the pattern of the rest of the data by including an indicator variable for the last timepoint.

If a separate measure of exposure within each time period (e.g. quarter) is available, such as the number of aircraft flying hours for a particular type of aircraft, or the number of takeoffs or landings, which might vary over time, this can be incorporated as an offset into the analysis, effectively modelling a rate, the number of occurrences per unit of exposure, as follows:

$$\ln(\mu_t) = \alpha + \beta t + \ln(\text{Exposure}_t)$$

$$\ln(\mu_t) - \ln(\text{Exposure}_t) = \alpha + \beta t$$

$$\ln(\mu_t \,/\, \text{Exposure}_t) = \alpha + \beta t$$

$$\text{Average number of occurrences per unit of exposure} = \frac{\mu_t}{\text{Exposure}_t} = \exp(\alpha + \beta t)$$

Modelling with the offset in the explanatory model permits the observed counts to still be modelled as a Poisson-distributed variable with average μ_t, whereas if we divided the observed counts by the exposure measure to create a rate before analysing, we would lose the Poisson distribution, and the relationship of the variance to the mean, whereby larger counts tend to have larger variability – whether those larger counts are due to increased exposure or increased rate of occurrences. Then the assumed statistical model would be inappropriate and it would be difficult to properly assess whether an individual value or pattern is unusual.

If in fact exposure in each time period did vary over the periods included in an analysis, incorporation of a relevant measure of exposure would facilitate a superior analysis, by explaining more of the observed variability and thereby allowing greater sensitivity to detect departures from the expected trends or patterns.

Appendix B. Summary of Models Fitted for Screening

The full modelling process as recommended by Data Analysis Australia is described here, containing more statistical modelling details. A sequence of models is fitted and nested models (where one model is a special case of another) are compared to determine statistically significant trends and patterns.

A general principle of statistical model-fitting is to choose the simplest or most parsimonious model (fewest parameters fitted) that adequately describes the variation in the data. We call this the best-fitting model.

Initially we choose the best-fitting model out of M1 to M4 (see Figure 3), and the flags for LastSpecial, Trend, Underdispersion and Overdispersion are based on these alone.

We then fit two models that generalise one of these (see Figure 11):

M5. Change in level – begins at one level, and at an estimated changepoint jumps (up or down) to another level. Three parameters must be estimated. Model M2 is a special case of this, with the changepoint set between the last two observations, and thus, in the lattice diagram, M5 sits above M2 with an arrow leading to M2.

M6. Quadratic trend model – enables a crude test for evidence of nonlinearity, by adding a squared term to the linear trend model M3.

If model M5 is found to be statistically significantly better than M2, a flag for Change in level is raised. If model M6 is statistically significantly better than M3 (the Q term is significant), then a flag for nonlinearity is raised.

Of course many other models are possible, but we chose to limit the investigation to models with up to three parameters to be estimated. With only 20 or 24 quarterly observations (from 5 or 6 years) estimating more parameters could be misleading.

These six models cover a range of possible shapes with the opportunity to identify different ways to describe the trends or patterns and to detect possible safety issues.

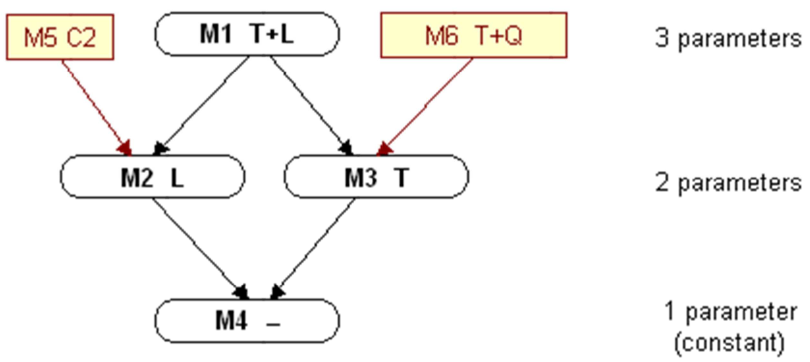


Figure 11. The models of Figure 3 (Section 4.1.1) with the addition of models M5, Change in level at an estimated changepoint, and M6, Quadratic trend. T indicates linear Trend, L indicates Last observation special, C2 indicates Change in level at an estimated Changepoint (two parameters to be estimated), and Q indicates a Quadratic trend term.

Finally, notwithstanding the comments about the number of parameters fitted, we examine possible annual seasonal cycles in a limited set of models, M1 to M4. A simple sinusoidal cyclic seasonal term requires two parameters, which can be thought of as an amplitude (how large the oscillations are) and a phase shift (where the peak occurs along the cycle). (In fact we fit them as the sum of a sine and a cosine curve – a mathematically equivalent parameterisation but simpler to fit.) This yields models M7 to M10 in Figure 12.

The Seasonality flag is triggered if an acceptable model with seasonality (one of M7 to M10) is significantly different from the corresponding model without seasonality (among M1 to M4) – this is a two degree of freedom test because it involves dropping two parameters simultaneously – it is meaningless to consider the amplitude or the phase in isolation – unlike the other comparisons which are all tests with one degree of freedom.

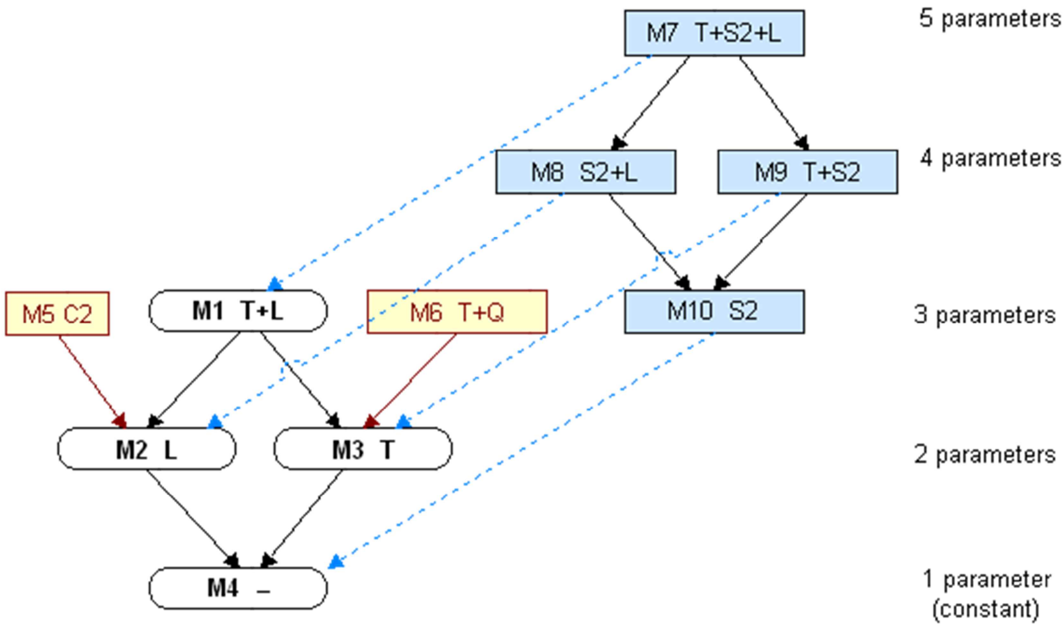


Figure 12. Figure 11 with the addition of annual seasonal models. C2 indicates a change in level at an estimated Changepoint (two parameters), T indicates (linear) Trend, Q indicates Quadratic trend, S2 indicates Seasonality (two parameters), and L indicates Last observation special.

Figure 12 displays all the models considered in the suggested approach. The number of parameters in each model (the same within a row) is shown on the right hand side.

Every model includes a constant or intercept parameter. Each additional code letter indicates one added parameter, except for C2 and S2 which each require two parameters.

Following an arrow from any model to a model lower in the lattice corresponds to dropping a term from the model, and the second model is a special case of the first. We say the second model is *nested within* the first. A statistical hypothesis test comparing these two models is a test of whether that dropped term is needed. If the test is statistically significant, that term is important and cannot be dropped.

Whan a Poisson distribution is assumed, the tests are chi-square tests with degrees of freedom corresponding to the difference in the number of parameters estimated – thus one degrees of freedom for most tests, but two degrees of freedom for the tests of seasonality, corresponding to the blue dashed lines in Figure 12.

Note that the statistical significance of any given term depends on which other terms are in the model. For example a linear Trend term might be not significant if a Last special term is present (comparing models M1 and M2), but significant if the Last special term is absent (comparing models M3 and M4).

Typically, we follow a path down through the lattice, testing at each step, dropping terms that are not statistically significant, until we can simplify the model no further. The aim is to find the simplest model (most parsimonious, with fewest parameters) that adequately describes the data.

A single unambiguous "best" model is not guaranteed. For example we may find that M1 and and M2 are not significantly different, but we cannot proceed to M4, and also that M1 and M3 are not significantly different, but we cannot proceed from M3 to M4. Either model M2 or model M3 might be acceptable – we need *either* Trend or Last special to explain the variation in the data, but *not both* – we don't need the generality of model M1. In this case we choose between models M2 and M3 on the basis of a measure of goodness of fit for each model, but we would note the ambiguity.

In the procedure recommended by Data Analysis Australia for the CASA safety incident screening, we focus initially on selecting the best model from among models M1 to M4. Then we consider some generalisations of these models (M5 and M6) to flag possible different causes of variation, and finally we consider the effects of annual seasonal cycles (models M7 to M10).

Dispersion

By preference, these models for count data are fitted using the Poisson distribution and tests comparing nested models (one model is a special case of the other, connected by a line in Figure 12) are based on the chi-square distribution for the change in deviance. This is a likelihood ratio test. However, if overdispersion is identified in model M1, this is flagged, and the quasi-likelihood method is used.

Dispersion is tested in model M1 using the Pearson chi-square statistic, sum of $(\text{observed} - \text{expected})^2/\text{expected}$, with degrees of freedom equal to the residual degrees of freedom from model M1.

If dispersion is found to be statistically significantly greater than 1, this is flagged as overdispersion and the quasi-likelihood method is used, specifying the distribution family as quasi-Poisson. Dispersion is estimated as the Pearson chi-square statistic divided by its degrees of freedom.

Then the likelihood ratio chi-square tests are replaced by F tests, with the F statistic calculated as the change in deviance divided by the change in degrees of freedom, divided by the estimated dispersion, and compared with an F distribution with numerator degrees of freedom equal to the change in degrees of freedom (which is the same as the change in the number of parameters estimated) and denominator degrees of freedom equal to the residual degrees of freedom in the more general model (which is the same as the number of observations minus the number of parameters estimated in this model).

If the dispersion is statistically significantly less than 1, this is flagged as underdispersion, and suggests unusually small variation, possibly indicating improper recording of incidents and worthy of investigation. However the ordinary Poisson model is fitted.

Flags

The circumstances under which each flag will be triggered will be clearly defined based on the sequences of models fitted as displayed in Figure 12, and other procedures.

Appendix C. Example Summary Display of Outcomes

This Appendix shows an example of the type of summary display that could be presented in an Excel spreadsheet to indicate the various anomalies, patterns or changes in the data flagged by the analysis. See Figure 13. There is one row for each occurrence type, and a column for each flag that could be raised, indicating different aspects of the pattern or variability of the numbers of incidents for that occurrence type.

The columns are ordered roughly in order of importance, with the earlier columns more likely to indicate aspects of concern or interest in relation to aviation safety, and the later columns more likely to provide background information. Together they provide a concise but comprehensive picture of characteristics of the data that might have relevance to aviation safety, and information to help determine what might have caused those patterns.

A graphical plot of the data and the most relevant fitted models for each occurrence type is an invaluable supplement to this table, and it is expected that such graphs will be referred to for any occurrence types that warrant further investigation. It might be possible to include a clickable link to the relevant graph for each occurrence type in the table.

The columns of the table and the meaning of the symbols are described below.

The first three columns contain identifying and background information:

- **Occurrence Type**. Could be individual occurrence types, or groups of occurrence types, or data grouped by geographical region, etc.
- **Average Count**. Average number of occurrences per quarter over the period of this analysis (e.g. four or five years). Included for background information.
- **Last Count**. Number of occurrences in the last quarter in this analysis. Included for background information.

The next eight columns are flags based only on models M1 to M4, considering one or both of a linear trend and a special term indicating the last observation is different, or a constant mean across the whole time period:

- **Last special**. The last observation deviates from the pattern (linear trend or constant) for the prior observations. The direction and strength of evidence is indicated by bold (🡅, 🡇) or finer (↑, ↓) arrows, and colour-coded as in the CASA START tool to strongly highlight this most important indicator of something different happening in the last quarter.
- **Trend**. A linear trend is evident in the data. A sloping arrow (↗ or ↘) indicates the direction.
- **Second-last special**. When the last observation is omitted, the second-last observation deviates from the prior pattern. Indicated similarly to the Last special flag (🡅, ↑, ↓ or 🡇). Outcome might differ from outcome of the previous run due to using slightly different data.

- **Last quarter changes pattern**. This flag will be triggered if, for example, there is no linear trend prior to the last quarter, but when the last quarter is included a linear trend becomes evident, or vice versa. Indicated by # (to indicate change in *pattern*), possibly with a very brief indication of which aspect of the pattern changed.
- **Underdispersion**. Displayed as ^^^, this flag indicates that, in the model including linear Trend and Last special, there was less variability in the data than expected from a Poisson distribution – the counts were unusually consistent. A *possible* cause of this is inadequate reporting, and therefore this might warrant checks on the adequacy of reporting. Poisson models are still fitted for all the screening tests, despite the underdispersion.
- **Overdispersion**. Displayed as /\/\/\, this flag indicates that, in the model including linear Trend and Last special, there was greater variability in the data than expected from a Poisson distribution – there is unexplained extra variation. This is analogous to the high variability flag reported in in the CASA START spreadsheet. In the models fitted and statistical tests used, a quasi-Poisson model is used to adjust for this.
- **Excess zeros**. This flag triggers if the number of zeros in the data is unusually large for a Poisson distribution, and is displayed as 000. This could be evidence of under-reporting in some of the quarters. It is tested by fitting a zero-inflated Poisson version of model M1 and examining the statistical significance of the term for the zero inflation. Zero-inflation may be one explanation for overdispersion, but we cannot say if it is the sole cause.
- **Outliers**. If the data contains one or more gross outliers, this can affect other tests. For example a single outlier can increase the variation such that trends appear not significant and therefore would not be flagged. Therefore if this flag triggers, the reader should be aware that other characteristics of the data might be obscured. A + indicates unusually large values, while a – indicates unusually small values, based on residuals from model M1. In an individualised analysis, the models would be re-examined with the outliers omitted, but that is less feasible in an automated procedure.

The next two flags are based on two generalisations of these four basic models:

- **Change in level**. Generalises the Last special model to allow a change in level at any time point. The flag triggers if the Change in level model is statistically significantly better than the Last special model. Displayed as a step symbol, indicating either a step up (_┌─) or a step down (─┐_). The quarter in which the change occurred could also be shown.
- **Nonlinearity**. Generalises the linear Trend model by adding a quadratic term to provide a crude test for nonlinearity. The flag triggers if the quadratic term is statistically significant. The symbol displayed, ◡ or ◠, indicates whether the departure from linearity tends to be upwards or downwards.

There are three more flags:

- **Seasonality**. Models M1 to M4 are fitted with the addition of a simple annual cycle term to investigate seasonality. This flag triggers if the seasonality is statistically significant, and the symbol ~~ is displayed to indicate the presence of an annual cycle.
- **Dispersion changed**. Overdispersion indicates extra variation from unexplained sources. Sometimes that variation can be explained by including other terms such as quadratic trend or seasonality, and the apparent overdispersion disappears in these models. This is indicated by the symbol ~~ΛΛΛ~~. In other cases, the additional terms might explain the data so well that underdispersion appears – possibly indicating over-fitting which means the model might be unrealistic and cannot be relied upon. This is indicated by the symbol ^^^ in this flag. The model term(s) associated with the change in dispersion can also be shown, for example _┌─, ◡ or ~~. Overdispersion might also be due to zero-inflation, but we cannot say if that is the sole cause, so that will not appear here.
- **Trend over time in rate per unit exposure**. This flag is available only if a measure of exposure has been used as an offset in the model. Currently, such measures are not available. If, after allowing for exposure, a (linear) trend is evident in the data, this flag will trigger. This indicates that the number of incidents per unit exposure (a rate) has changed over time. An upward or downward sloping arrow indicates the direction.

The symbols used for each flag hint pictorially at the characteristic flagged, to aid quick interpretation. In addition, a plot of the data and some relevant fitted models will guide the reader. For compactness of this table, these would be provided separately, possibly with a clickable link from this table. However, the table itself indicates the main features. For example, for the Large counts example in Figure 13, the flags suggest the last observation stands out larger than expected from an otherwise general downward trend, and the variation generally is larger than expected for Poisson counts (overdispersion). This can be compared against the data and fitted models displayed in Figure 4 in Section 4.6.2.

Occurrence Type	Average Count	Last Count	Last Special	Trend	Second-Last Special	Last Quarter Changes Pattern	Under-disp-ersion	Over-disp-ersion	Excess Zeros	Outliers	Change In Level	Nonlin-earity	Season-ality	Dispersion changed	Trend in rate per unit exposure
Incident category 1			**↑**	↗	**↑**		^^^		000						
Incident category 2			↑	↘	↑			/\/\/\		+	_┌─	◡		~~^^^~~ _┌─	
Incident category 3						# trend				−	─┐_	◠	~~		↗
Incident category 4			↓		↓				000						↘
Incident category 5			**↓**		**↓**			/\/\/\						~~^^^~~ ~~	
…															
Controlled.airspace	218.5	207		↘				/\/\/\							
Verbal.instruction	275.7	309		↗				/\/\/\							
Overweight.landing	0.35	0				# trend						◠			
Passenger.related	4.5	4						/\/\/\							
Runway.Incursion	73.2	67						/\/\/\		+		◠			
Airframe.overspeed	7.1	9						/\/\/\	000		_┌─	◡			
Artificial data:															
Large counts example	218.5	270	**↑**	↘				/\/\/\							
Small counts example	0.35	2	**↑**	↗		# lin, quad						◠			
Strong linear example	121.5	62	**↓**	↘			^^^								
Seasonality example	8.2	12	↓	↗							_┌─	◡	~~	^^^ ~~	
Zero-inflated example	2.6	12	**↑**					/\/\/\	000						
…															

Figure 13. Example of Excel spreadsheet summarising screening results and flags to inform further investigation.