Nationwide Indicators of Surface Safety at Towered Airports

Jacques Press

July 2007

DOT/FAA/TC-07/23

U.S. Department of Transportation
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

NOTICE

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full Text Technical Reports page: http://actlibrary.tc.faa.gov in Adobe Acrobat portable document format (PDF) form.

The author can be reached at jacques.press@faa.gov

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/		

4. Title and Subtitle	5. Report Date / 6. Performing Organization Code
Nationwide Indicators of Surface Safety at Towered Airports	5. Report Date: July 2007 6. Performing Organization Code: AJP-1730
7. Author(s) Jacques Press	8. Performing Organization Report No.: DOT/FAA/TC-07/23
9. Performing Organization Name and Address Federal Aviation Administration Systems Engineering Support Services Group William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration National Runway Safety Program Office and Safety Risk Reduction Information Group Office of Safety Services (AJS-1) Federal Aviation Administration 800 Independence Ave., SW Washington, DC 20591	13. Type of Report and Period Covered Technical Note 14. Sponsoring Agency Code AJS-4 and AJS-7

15. Supplementary Notes

16. Abstract

This document introduces a time-based method to aggregate airport incident information into nationwide indicators of safety. The objective is to revise how incidents are summarized conventionally today. Although the typical towered airport serves as context here, the idea is general enough to extend to facilities where incident numbers are used to measure safety performance. The document promotes time as a key variable in the incident tracking process. Accordingly, it recommends *days between incidents* as a measure of choice, besides the incident counts themselves. Extensive analysis of the Federal Aviation Administration Runway Safety Database from Fiscal Year 2000 to 2006 eventually points to three indicators. They are the monthly (a) incident count collected from each airport, (b) median number of days between incidents for each airport, and (c) percentage of incident considered serious. Collectively, the three are taken to portray the state of safety based on incidents in the national airport system.

17. Key Words	18. Distribution Statement
Safety, incidents, airports, indicators, control charts	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	51	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Page left intentionally blank

Table of contents

Page

1.0 Summary 11

1.1 Purpose 11

1.2 Incident recurrence 11

1.3 Three proposed indicators 11

2.0 Recommendations 13

2.1 Incident tracking 13

2.2 Perspective 13

2.3 Tracking incidents 13

2.4 The positive image of safety 14

3.0 Introduction 15

3.1 National measures of surface safety 15

3.2 Motivation 16

3.3 A safety measurement framework 16

4.0 Safety View 17

5.0 Data Source 18

6.0 Units of Measure 19

6.1 Incidents 19

6.1.1 Incident types 19

6.1.2 Incident sources 20

6.1.3 Incident frequency by type and source 20

6.2 Time between incidents 21

6.2.1 Sample time computation 21

6.2.2 Special cases 22

6.2.3 Frequency distribution 23

6.2.4 Summary statistics 23

6.2.5 Mean, median, skewness 24

6.2.6 Airport safety performance statistics 24

Table of contents (cont.)

Page

7.0 Methodology 28

7.1 Survey 28

7.2 Selection criterion 30

7.3 Selection process 30

7.4 Aggregation and display 32

7.4.1 Occurrence 32

7.4.1.1 Aggregate occurrence by local hour of the day 32

7.4.1.2 Aggregate occurrence by month of the year 33

7.4.1.3 Occurrence control charts 34

7.4.1.4 Special causes 35

7.4.2 Recurrence 37

7.4.2.1 Aggregation 37

7.4.2.2 Mean (average) and median times between incidents 38

7.4.2.3 Recurrence control charts 41

7.4.3 Seriousness 43

7.5 Statistical summary 44

Appendix A: Definitions 48

List of Figures

Page

Figure 1.1 Incident occurrence, recurrence and seriousness as three nationwide aggregate safety indicators 12

Figure 3.1 A safety measurement framework 17

Figure 5.1 Airport incident trends by fiscal year 18

Figure 6.1 Distribution of incidents by type.............................. 19

Figure 6.2 Distribution of incidents by source 20

Figure 6.3 Days between a pair of incidents at hypothetical Airport DEF... 21

Figure 6.4 Single incident at hypothetical Airport GHI 22

Figure 6.5 Same-day pair of incidents at hypothetical Airport JKL 22

Figure 6.6 Frequency distribution of "days between incidents" over seven years 23

Figure 6.7 Airports listed by incident count and average number of days between incidents over seven years 27

Figure 7.1 Methodology 28

Figure 7.2 Selection process 31

Figure 7.3 Airport ABC incidents aggregated by local hour of the day 32

Figure 7.4 Airport DEF incidents aggregated by local hour of the day 33

Figure 7.5 Nationwide incidents aggregated by local hour of the day 33

Figure 7.6 Nationwide incidents aggregated by month of the year... 34

Figure 7.7 Incident occurrence as a nationwide aggregate safety indicator..................................... 36

List of Figures (cont.)

Page

Figure 7.8 Moving range for incident occurrence 36

Figure 7.9 The incident recurrence aggregation process.......... 37

Figure 7.10 Computation of average days between incidents... 39

Figure 7.11 Boxplots of monthly distribution of median days between incidents 40

Figure 7.12 Incident recurrence as a nationwide aggregate safety indicator..................................... 42

Figure 7.13 Moving range for incident recurrence..................... 42

Figure 7.14 Incident seriousness as a nationwide aggregate safety indicator..................................... 43

Figure 7.15 Frequency distribution of the three indicators 45

List of Tables

	Page
Table 1.1 Three nationwide aggregate safety indicators	11
Table 3.1 Towered airports across the United States	15
Table 3.2 Types of airport surface incidents	15
Table 6.1 Incidents crosstabulated by type and source	21
Table 6.2 Days between sequential incidents at Airport XYZ	22
Table 6.3 Statistical results for days between incidents (measures of data location) for pooled data	25
Table 6.4 Statistical results for days between incidents (measures of data dispersion) for pooled data	26
Table 7.1 Global/national measures in several domains	29
Table 7.2 Selection criterion	30
Table 7.3 Results of the selection process	31
Table 7.4 Special causes for the *occurrence* indicator	35
Table 7.5 Sample monthly list of airports, incident counts mean (average) and median days between incidents for December 2001	38
Table 7.6 Special causes for the *recurrence* indicator	41
Table 7.6 Special causes for the *seriousness* indicator	43
Table 7.7 Statistical results (measures of data location)	46
Table 7.8 Statistical results (measures of data dispersion)	47

Page left intentionally blank

1.0 Summary

1.1 Purpose

This document introduces a time-based method to aggregate airport incident information into nationwide indicators of safety[1]. The aim is to revise how incidents are summarized conventionally at the Federal Aviation Administration (FAA). Although the typical towered airport serves as context here, the proposed approach is general enough to extend to other types of facilities where incident counts are used as measure of safety performance.

1.2 Incident recurrence

Despite its revisionary intention, the approach remains incident-centered, endorsing the belief that incidents reflect the state of safety, until such doctrine is replaced by a better one. Yet, one of the ideas proposed here goes a step further. It introduces time as key variable in the incident tracking process. For instance, it emphasizes *days between incidents* as an important measure, besides the incidents themselves.

Labeled *recurrence*, this time-based measure conveys a favorable, positive image of safety by expressing how long the facility remains incident-free. Recurrence digresses from traditional incident counting which, by its inherent definition, supplies the unfavorable, negative safety view. Recurrence carries much parallel with the practice of charting days without injury in the workplace. Within that same spirit, it is singled out to be aggregated for all airports. This course of action leads to a positive safety indicator suitable for the country as a whole, once this measure is launched into practice. By way of illustration, the *recurrence* control chart[2] in the middle of Figure 1.1 shows increasingly longer times between incidents, a marked safety improvement during the last seven years.

1.3 Three proposed indicators

While recurrence is being underscored, the two conventional measures, incident counts and severity, remain important, not to be dismissed. Labeled *occurrence* and *seriousness* here, both should be given a wider meaning by expressing them as functions of time, like in Figure 1.1, if they are to carry any substantial inferential value. Without this time element woven in their fabric, little can be generalized (estimated, predicted) about their trends. In summary, all three indicators are defined below in Table 1.1. Collectively, they are taken to portray the state of surface safety in the national airport system.

Indicator*	**Definition**
occurrence	incident count collected from each airport
recurrence	median number of days between incidents for each airport
seriousness	percentage of incident considered serious

*aggregated monthly from nationwide data

Table 1.1 Three nationwide aggregate safety indicators

[1] See Appendix A for definitions of terms.
[2] The Statistical Package for the Social Sciences (SPSS) is used to generate most of the tables and all the plots in this document.

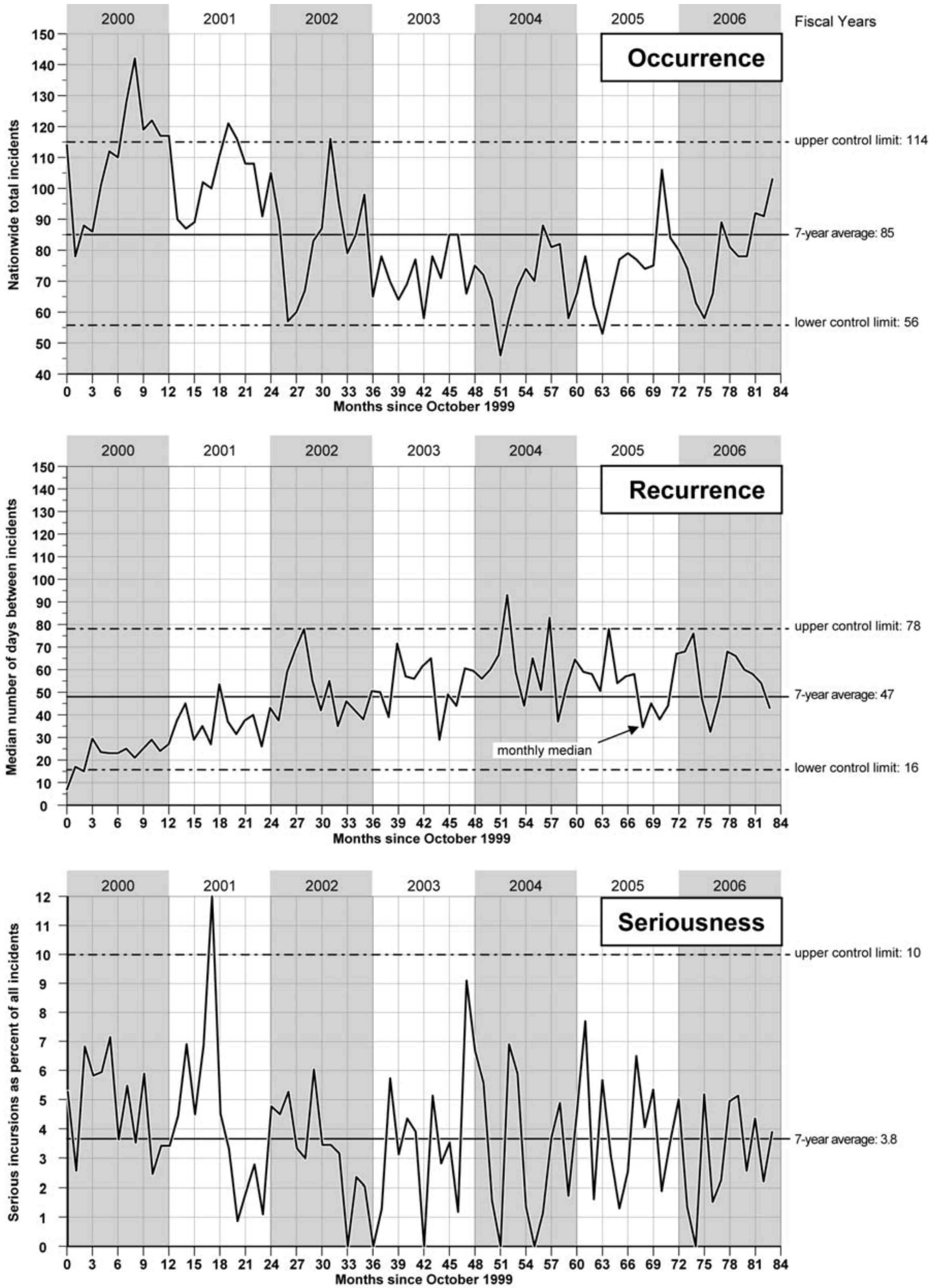


Figure 1.1 Incident occurrence, recurrence and seriousness as three nationwide aggregate safety indicators*

*based on 7,129 surface incidents at 466 towered airports, from October 1999 to September 2006.

2.0 Recommendations

2.1 Incident tracking

Charts like those in Figure 1.1 should be part of the FAA's practice to track incidents statistics. Moreover, this kind of chart belongs in the agency's yearly *Flight Plan*[3] document to highlight how safety is improving nationwide on the airport surface. Without these charts as tools, today's tracking continues to resort to counting aimed at simply comparing differences month-to-month or year-to-year. This limited process derives most likely from the belief that sensitive information must be simple and clear-cut to be accurate. However, despite their attractive simplicity, count differences alone do not bring deeper insight. Little, if any analytical inference stems beyond them.

2.2 Perspective

To avoid misjudging safety changes implied by incident counts alone, the aviation community should find out the *wider* meaning of these numbers. For example, safety audiences are entitled to know what 15 over 12 incidents a year ago imply for Airport XYZ. A right perspective helps in this process. Perspective comes naturally if the reported numbers are sized relative to a more detailed calendar time and to the hundreds of airports servicing the nation. Towards that goal, much needed on the agenda is an analytical approach, as described in this document, supplementing the enumerative way of reporting surface safety. Beyond today's restrictive formats, this new approach would encourage inferential trending and forecasting by relying on time-based control charts, as exemplified on the previous page.

2.3 Tracking incidents

On an even broader scale, control charts should be used regularly by those commissioned to chronicle aviation incidents, be they airport-centered or otherwise. Control charts are not simply attractive graphical tools. Instead, they are derivatives of a larger, underlying *quality assurance* process. That process, conceived by Shewhart and later amplified by Deming[4], enlists formal rules to isolate out-of-control trends, the analytical rather than intuitive way. This recommendation rejoins the safety thinking of several sources, like Stolzer et al[5], Batson et al[6], Cutler[7], Salazar[8] , Cheng et al[9] and Sorrell[10]. In particular, Stolzer and Cutler have

---

[3] See the latest Federal Aviation Administration Flight Plan, for example.

[4] Deming, W. E. , (1982). *Out of the Crisis: Quality, Productivity and Competitive Position*, Cambridge Univ. Press.

[5] Stolzer, A. J. Wu, H., Halford, C. (2006). "Six Sigma Applied to Flight Operations Quality Assurance: An Examplar Case Study", International Journal of Applied Aviation Studies, Vol. 6, No. 1, pp. 11-24.

[6] Batson, R. G., Moynihan, G. P, Zhou, L., (2000). " Control Charts for System Safety Indicators", American Society for Quality (ASQ), Vol. 54, No. 0, pp. 177-189.

[7] Cutler, A. N., (1999). "Normal Accidents: A Statistical Interpretation, Royal Statistical Society Conference on Risk, Warwick, UK.

[8] Salazar, N. (1989). "Applying the Deming Philosophy to the Safety System", Professional Safety, 34(12), 22-27.

[9] Cheng, A. Y., Liu, R. Y., Luxhoj, J. T., (2000). "Monitoring Multivariate Aviation Safety Data by Data Depth: Control Charts and Threshold Systems", IEE Transactions on Operational Engineering, Vol. 32, No. 9, pp. 861-872.

already considered how directly the popular Shewhart-Deming theory of quality control can be applied to aviation safety tracking. As an operational example among public safety agencies, the U.S. Coast Guard started using control charts back around 1999 to track casualties in the high-risk fishing ship industry[11].

2.4 The positive image of safety

In a complex facility like an airport, incidents are inevitable regardless of severity. History tells us that absolute safety eludes control as soon as the facility is put into motion. Aware of this reality, safety-sensitive groups are seen searching constantly for creative doctrines to safeguard their systems. A modern one involves adopting the non-punitive *just culture* where no particular individual is penalized for mishaps[12]. In that context, the recurrence idea, presented here as incident-free intervals, offers a productive, non-negative alternative to those favoring this new safety philosophy.

[10] Sorrell, L., (1998). "Safety and Statistical Process Control: One Practitioner's Perspective", Professional Safety, Vol. 43, Issue 1, page 37, Editorial.

[11] Spitzer, J.D. (1999). "Report of the Fishing Vessel Casualty Task Force", U.S. Coast Guard, DOT.

[12] Gordon, R., Moylan, P., Stastny, P., (2004). "Improving Safety Through a Just Culture", Flight Ops/ATC Ops. Safety Information Sharing, Global Aviation Information Network, Working Group E, Journal of ATC, October-December 2004, pp. 34-37.

3.0 Introduction

3.1 National measures of surface safety

Given the public importance of a reliable airport system, how safely it delivers its services concerns constantly users and regulators alike. To address it, one desirable action would be to identify national measures which can synthesize safety information from many airports in an explicit yet compact sense for all to understand. Only then can they become *barometers* of safety answering the perennial self-question "how well are we doing safety-wise at the nation's airports?"

Searching for compelling answers, one might consider intuitively accident and incident records as candidate measures. However, summarizing them into a few nationwide indicators is a challenging undertaking because there are hundreds of airports and several kinds of incident possible on their surface. Tables 3.1 and 3.2 give an idea of these high numbers.

Airport type	Count	Provided services
Federal towered	263	FAA Air Traffic Control (ATC) services
	231	Contractor-provided ATC services
Federal subtotal:	494	
Military towered	164	ATC services provided by military air traffic controllers

Table 3.1. Towered airports across the United States[13]
(source: FAA, "A Plan for the Future 2006-2015", June 2006, page 15)

Incident Type[14]	Count[15]
Operational deviations (OD)	128
Operational errors (OE)	654
Pilot deviations (PD)	3,892
Vehicle/pedestrian deviations (V/PD)	2,455
Total	7,129

Table 3.2. Types of airport surface incidents

Loss of representation is the main drawback in generalizing safety from many sources into a limited set of iconic expressions. Therefore, as a precaution, a rigorous approach to be presented shortly must eventually formalize any intuitive start in defining conglomerate measures.

[13] FAA, "A Plan for the Future 2006-2015", June 2006, page 15.
[14] See Appendix A for definitions
[15] Source: FAA, Runway Safety Database, Office of Safety Services, Fiscal Year 2000 to 2006.

3.2 Motivation

Conglomerate measures, particularly those promoted as official metrics, are not new to aviation. Many of them find useful life in important sub sectors, including air traffic control. They range from simple to complex. Consider for instance *delay reduction per air traffic system cost increase* and *restricted airspace availability days per month multiplied by weeks of advance notice.*[16] Unlike these capacity metrics, safety-oriented ones still await consensus. As late as 2006, Brooker[17] questions how well safety measures can come into use. And, while air traffic delay figures are being aggregated into metrics as baseline for capacity planning, no similar quantitative benchmark is yet calculated for the safety side. Yet, such benchmark would help assess the safety impact of future functionalities like those proposed in new generations of air traffic control systems.

Concern for the future is not the only motivation here. On a wider scale, universal measures of safety have been desirable in the aviation community for a long time. The literature points to several past approaches competing to generate superlative measures. Conceived over the years, most of them involve rates whose numerators and denominators represent intuitive variables like fatalities, accidents, etc. Other variables are hours and miles flown between accidents. Think, for example, of the traditional one: *number of accidents per 100 million passenger miles*. Several metrics like this one have been summarized decades ago by Villareal[18] along with their advantages and disadvantages in expressing the state of aviation safety.

3.3 A safety measurement framework

In fashion similar to existing aviation metrics, the objective is to find measures qualifying airport surface safety with sparkling clarity. Conditional to any finding, they must not only be measurable but also representative of the nation as a whole, once they are aggregated across all airports. To meet that difficult objective, several starting premises are in order. Together, they organize a safety measurement framework[19] formalizing how to proceed in a rational way. This precaution is necessary because without framework, the search can turn easily haphazard and fruitless. Therefore, four framework premises are chosen here as core principles, assuming they match sufficiently well the general intention expressed in this document. Sketched in Figure 3.1, they are in logical order:

- a view of safety consistent with our end purpose,
- a data source providing supporting evidence for that view,
- units of measure acceptable as "common currency" for airport safety, and
- a method to identify, aggregate and display these units as indicators.

[16] For an extensive overview see Report of the Air Traffic Services Performance Focus Group, (1999). "Airline Metric Concepts for Evaluating Air Traffic Service Performance", CNS/ATM Focused Team.
[17] Brooker, P. , (2006). Cranfield University, "Are There Good Air Traffic Management Safety Indicators for Very Safe Systems?, Journal of Navigation, 60 (1), pp. 45-67.
[18] Villareal, C. T., (1988). "Uses and Misuses of Risk Metrics in Air Transportation", *Transportation Research Record*, 1161, pages 31-42.
[19] A framework is a mental model "framing" the targeted concept in terms of scope, assumptions, premises, entity-attribute relationships, order of precedence, etc.

Each premise is addressed in detail in the remainder of this document.

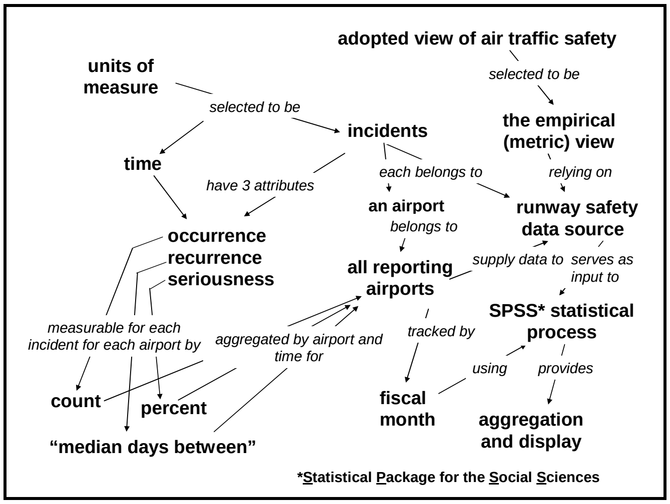


Figure 3.1 A safety measurement framework

4.0 Safety View

There are multiple ways to view the abstract notion of air traffic safety. Formal ones (like most separation models[20]) impose usually mathematical theory on their audience. If theory proves difficult to accept, one might switch to a more direct, empirical (data-based) view. It employs statistical mining tools to examine information collected during real time operations. The strategy is to discover patterns and relationships in the data without invoking any prerequisite theory. Through this heuristic approach, data mining results might invite the thought of generating empirical measures (*metrics*) as quantitative sentinels of safety[21]. According to this scenario, metrics do not have to match any preconceived mathematical theory. Instead they must make rational, practical sense. Such empirical image sounds attractive mainly because it relies on actual operational data to preserve realism in subsequent analyses. Given this advantage, it is elected to be part of the above framework.

[20] See for example, "A Concept Paper for Separation Safety Modeling", An FAA/EUROCONTROL Cooperative Effort on Air Traffic Modeling for Separation Standards, 20 May 1998.
[21] Press, J. (2003). "A Measurement Framework for Air Traffic Safety Metrics", Federal Aviation Administration, William J. Hughes Technical Center, DOT/FAA/CT-TN04/10.

5.0 Data Source

With the empirical view adopted, measures of all sorts can be made to capitalize on data sources maintained by the agency governing air traffic operations. For that purpose, today's officials keep thousands of records on defects, errors, incursions, violations, near-misses, incidents and, of course, accidents. All this information motivates naturally its custodians to analyze it in a number of different ways. Potential knowledge gained through data reduction and analysis is often the traditional justification why so much data is being warehoused in the first place.

Within the airport context, a particular source stands out as quite large. It is the runway safety database maintained at the FAA. It contains thousands of surface incident records (runway incursions and other surface events) spanning several years (starting with Fiscal Year 2000) and originating at hundreds of towers. Far from being a trivial list of events, this database is significantly dense with multiple, useful fields specifying each incident's characteristics and circumstances. Therefore, it too is made to join the framework.

Summarized in Figure 5.1, the runway safety data concordant with the framework's empirical view contains 7,129 non-duplicate incident records collected between Fiscal Years 2000 and 2006. Records show a trend going down from the Year 2000 to 2004. However that same trend rises in 2005 and 2006, implying there is no steady decrease since the Year 2000, as would have been desirable.

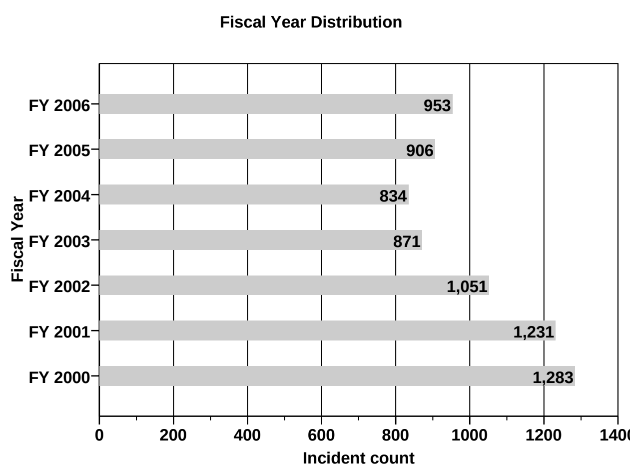


Figure 5.1 Airport incident trends by fiscal year

As caution, dependence on data can be precarious because no information remains complete and accurate at all times. However, the metric process to be adopted here does not require completeness. Instead, it claims only an estimate, subject to random sampling error bracketed by an upper and lower confidence limit, as expected in the traditional statistical sense. Such non-deterministic view is demonstrated in control charts later in this document. This approximation is further justifiable because not all unsafe events can be observed or reported anyway. Therefore, a complete enumeration is impossible, limiting any sort of measurement to just sampling. Despite this shortcoming, a sample should not be dismissed simply because it acts as surrogate to the entire population. To the contrary, a sample can be quite representative, once its size becomes sufficiently large, which is the case here.

6.0 Units of Measure

6.1 Incidents

Airport surface accidents are relatively few. As a result, their statistics would fall short when assessing safety trends. Incidents, or "near-accidents", are more numerous. Thus, they would be a next best choice in representing the manifestation of safety loss at the facilities which record them. Based on this belief, the analysis focuses on airport incidents as measures of safety. However, to be more complete, safety knowledge should also include the shape of their trend evolving over time.

6.1.1 Incident types

The term *incident* is defined here to mean any one of the following types of events:

- surface (non-incursion) incidents,
- serious runway incursions (severity A or B),
- non-serious runway incursions (severity C or D), and
- collisions.

See Appendix A for definitions of the above.

Summarized in Figure 6.1 below, the 7,129 incidents show a relatively small percentage (3.8%) tracing their lineage to serious runway incursions. The majority (65.7%) pertains to surface (non-incursion) incidents.

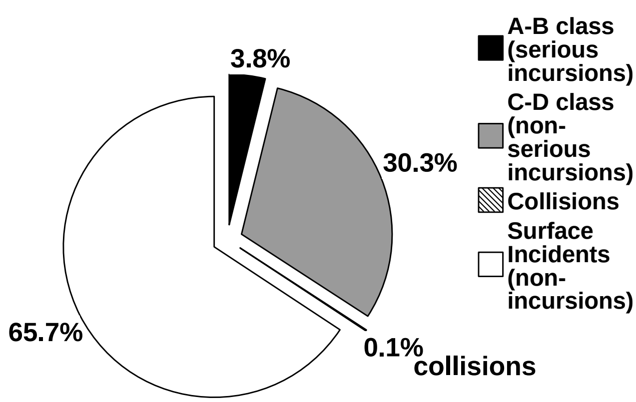


Figure 6.1 Distribution of incidents by type

As defined here, *incident* is assumed collective enough to mean all surface events, including runway incursions, even collisions. Also, note that incursions are being classified here in only two ways (serious or non-serious), replacing the traditional four letter codes (A, B, C or D) used to define severity. In this document's context, *serious* is sufficiently differentiable from *non-serious* because a nationwide aggregate measure (our end purpose here) should remain quite broad, not requiring any lower A, B, C, D classification. That is, we assume general audiences presented with this measure might ask only whether the incidents are serious or not, without further explanation. In other situations where a finer, more technical categorization is required, then letter codes should be used.

6.1.2 Incident sources

It is also assumed that incidents are caused by any one of the following sources:

- operational Error (OE) or deviation (OD) committed at the airport on the part of the air traffic controller,
- pilot deviation (PD) in taxiing, landing or taking off, and
- vehicle/pedestrian deviation (V/PD) at the airport surface.

See Appendix A for definition of these terms. Also see Figure 6.2 below where pilot deviations represent the largest source of incidents by far.

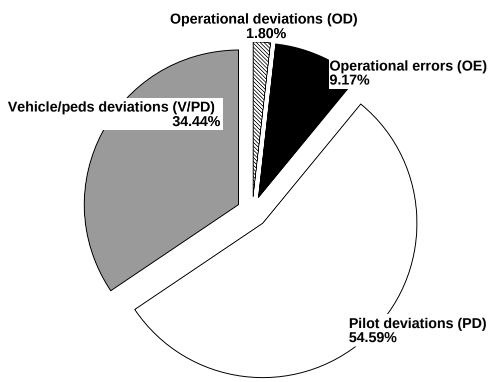


Figure 6.2 Distribution of incidents by source

6.1.3 Incident frequency by type and source

Adding more perspective, Table 6.1 shows how the same 7,129 incidents become distributed when their types are cross-tabulated with sources. Even here, pilot deviations carry a large share across incident types while operational errors appear mostly non-serious (505 in a total of 654).

Incident source	Incident type: A-B class (serious incursions)	C-D class non-serious incursions	Collisions	Surface Incidents (non-incursions)	Total
Vehicle/Pedestrian Deviations (V/PD)	39	410	1	2,005	2,455
Pilot Deviations (PD)	154	1,220	2	2,516	3,892
Operational Errors (OE)	81	505	5	63	654
Operational Deviations (OD)	0	27	0	101	128
Total	274	2,162	8	4,685	7,129

Table 6.1 Incidents crosstabulated by type and source

Absent from the framework are unobserved incidents which likely happen during operations. Also excluded are non-reporting airports. How much remains unknown is important. However estimating this unknown requires a separate analytical task, beyond the scope of this document. Hopefully, in the future, there will be enough real time sensor automation to capture all incidents at all facilities, regardless of the human capacity to do so.

6.2 Time between incidents

Besides counts, an important unit of measure deserves attention. It is the time interval between incidents, indicating how long the system remains incident-free until the next incident. Figure 6.3 illustrates this concept for hypothetical Airport DEF using the adopted unit of time "days between incidents" (dbi).

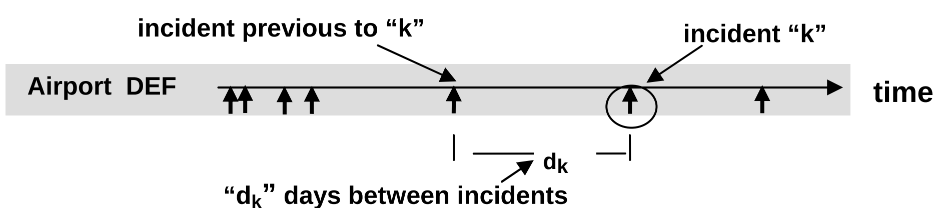


Figure 6.3 Days between a pair of incidents at hypothetical Airport DEF

6.2.1 Sample time computation

Consider a sample of incidents for Airport XYZ randomly selected from the data source and shown in Table 6.2. In that Table, Incident 1, being the first for that airport, carries no days between incidents. Incident 2 occurs 10 days beyond Incident

1 (see last column, second row in the Table), as computed using the shown calendar day/month/year difference between Incidents 2 and 1. Incident 3 occurs 10 days beyond 2, etc.

airport	incident sequence number	calendar day	calendar month	calendar year	days between incidents (dbi)
XYZ	1	1	7	2000	-
	2	11	7	2000	10
	3	21	7	2000	10
	4	9	8	2000	19
	5	17	8	2000	8

Table 6.2 Days between sequential incidents at Airport XYZ

6.2.2 Special cases

As computed in Table 6.2, a dbi involves a consecutive pair of incidents: the incident itself and its precedent at the same airport, where the smallest resolution for time between incidents is assumed to be the unit day. As a result, although Figure 6.3 represents the most common case data-wise, we include two extreme cases as follows. Extreme Case 1 applies when an airport experiences only one incident throughout the entire seven year period (2555 days), as shown in Figure 6.4. In that case, no incident pair is identified. Therefore, the incident counts as an occurrence but not as a recurrence.

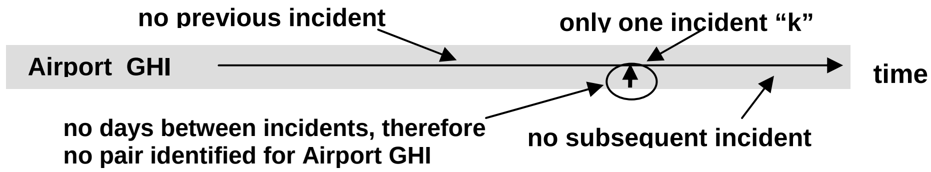


Figure 6.4 Single incident at hypothetical Airport GHI

Extreme Case 2 applies when an airport experiences two or more incidents the *same* day, as shown in Figure 6.5. In that case, incident pairs are created for them. Those pair are assigned zero dbi. Therefore, in Figure 6.5, we have two occurrences assigned with zero recurrence between them.

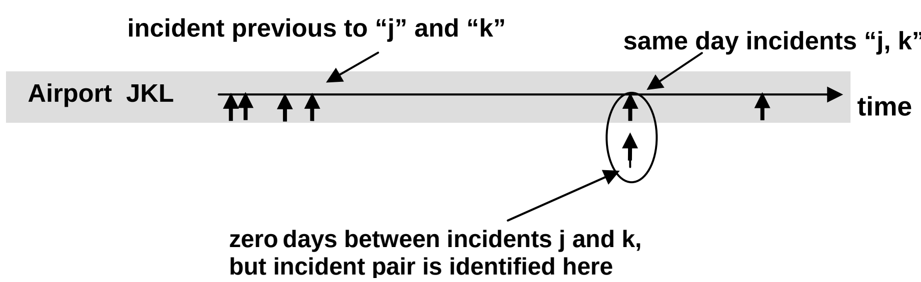


Figure 6.5 Same-day pair of incidents at hypothetical Airport JKL

6.2.3 Frequency distribution

Each of the 466 airports contributing to the nationwide sample must have one first-time incident. Therefore, the original 7,129 occurrences reduce to 6,663 recurrences (pairs, instances) expressed as days between incidents. The frequency of these pairs is shown in Figure 6.6, illustrating the variety of their time frames. There, the shape of the distribution agrees surprisingly well with the classical one used for modeling times between events[22].

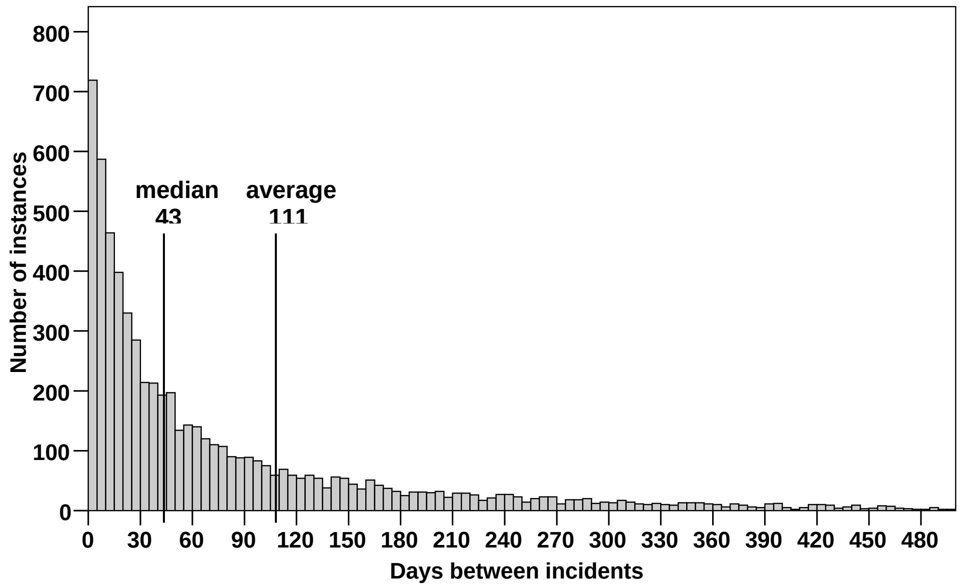


Figure 6.6 Frequency distribution of "days between incidents" over seven years

Like its theoretical counterpart, this distribution is "exponentially shaped". Only a few instances can be found beyond 500 days. Most of the intervals are short, about 30 days.

6.2.5 Summary statistics

The 6,663 instances are further summarized in terms of several statistics shown in Tables 6.3 and 6.4. Table 6.3 contains the traditional "measures of data location" (mean, median, etc.) It also contains other, more modern measures, like the 5% trimmed mean. The "Definitions" column in these two tables describes in words the displayed terms. Similarly, Table 6.4 provides "measures of data dispersion" (variance, standard deviation, etc.)

[22] See Montgomery, D. C., (2005). An Introduction to Statistical Quality Control. Wiley, 5h Edition, p.69.

6.2.6 Mean, median and skewness

As a central measure of the data, the mean (arithmetic average) in Figure 6.6 equals about 111 days. Unfortunately, such high, desirable value would be misleading because the data distribution is skewed, meaning it slides significantly to the right in the Figure. To remedy the situation, the median (43 days, about a third of the average) would be a better choice, being less sensitive to skewness. As confirmation, unlike the average, the median is closer to four other independent measures of data location (50.77, 37.47, 45.98, and 37.34 in Table 6.3) designed to be resistant to skewness.

6.2.7 Airport safety performance statistics

Also, the dbi's are inversely proportional to incident counts. This relationship is well confirmed in Figure 6.7. Furthermore, this Figure supplies a safety comparison between airports by serving as a 7-year overview of the nationwide performance, expressed as average days between incidents versus incident count at each facility.

Descriptive statistics for 6,663 recurrences over 7 years	**Definitions Units**	**days between incidents**
Mean (S.E)	arithmetic average and (standard error of the average)	110.64 (2.386)
95% Confidence Interval for the Mean	lower confidence limit	105.96
	upper confidence limit	115.31
5% Trimmed Mean	same as mean but computed with 5% of the lowest and highest ranked data values removed (trimmed)	79.21
Huber's M-Estimator	several resistant (robust) measures available in the Statistical Package for Social Sciences (SPSS). They are akin to the arithmetic average. However, they are more representative of the data[23], because each reflects a weighted mean less sensitive to outliers or skewness in the data. In general, values closer to the center of the data distribution are given more weight while distant values are given less weight.	50.77
Tukey's Biweight		37.47
Hampel's M-Estimator		45.98
Andrews' Wave		37.34
Median	the middle value in the data (50% percentile)	43.00

Table 6.3 Statistical results for days between incidents (measures of data location)

[23] Andrews, D.F., Bickel, P.J., Hampel, F.R., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, N.J.

Descriptive statistics for 6,663 recurrences over 7 years	**Definitions Units**	**days between incidents**
Variance	statistical measure of variation	37917.382
Standard deviation	square root of the variance	194.724
Minimum	lowest value	0
Maximum	highest value	2142
Range	highest minus lowest value	2142
Interquartile Range	same as the range but with the lowest and highest 25% of the ranked data values removed	109
Skewness[24]	measure of data distribution shift or "slide"	4.250
Kurtosis[25]	measure of data distribution peakedness or flatness	24.515

Table 6.4 Statistical results for days between incidents (measures of data dispersion)

[24] See Porkess, R. (2005). Collins Internet-linked Dictionary of Statistics, 2nd Edition.
[25] ibid.

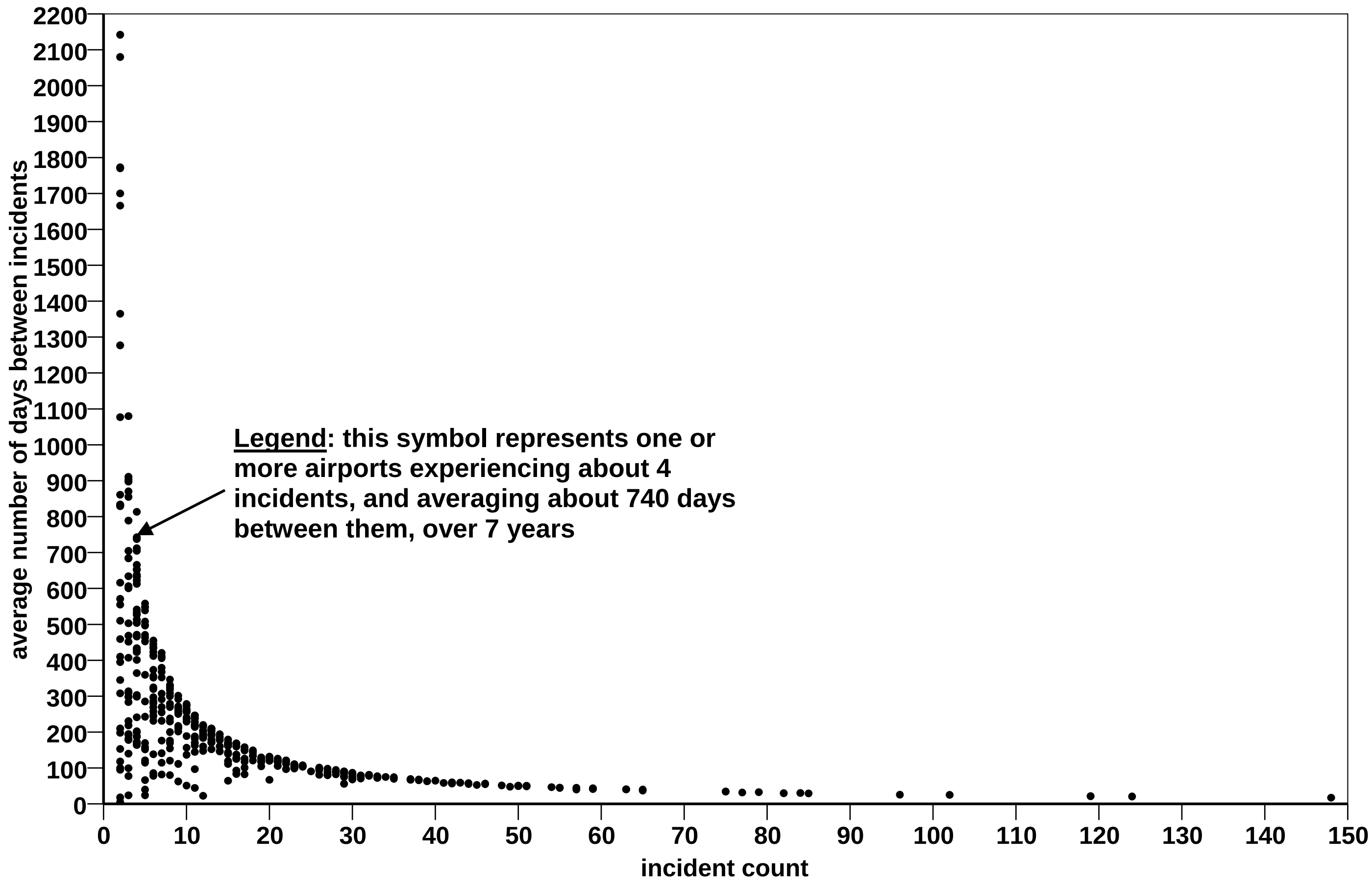


Figure 6.7 Airports listed by their individual incident count and average number of days between incidents over seven years

7.0 Methodology

According to the fourth and most important obligation to the framework (Section 3.0), a methodology remains to be devised to identify, aggregate and display indicators. Towards this end, several steps are outlined in the logical order depicted in Figure 7.1. They are addressed individually in the remainder of this section.

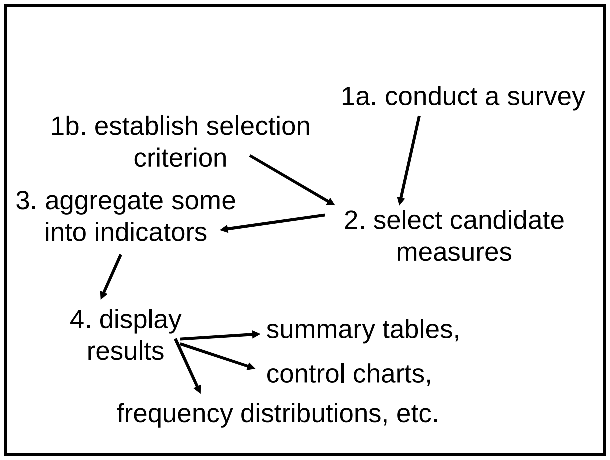


Figure 7.1 Methodology

7.1 Survey

For orientation purpose, it helps to examine how domains outside aviation derive their own measures of performance. Valuable methods and lessons learned elsewhere might be beneficial here. Accordingly, a survey conducted recently reveals the following two major observations:

(a) National, even global measures of performance (safety or otherwise) are widely present across several disciplines[26]. The disciplines fall into two major groups. One group represents high assurance, safety-sensitive domains, like law enforcement, health care and the environment. The other group includes non-safety domains like sociology, business (e.g. stock market) and economics. Table 7.1 below summarizes some of the findings.

[26] See for example Government Accountability Office (GAO) (2003). "Forum on Key National Indicators: Assessing the Nation's Position and Progress", GAO-03-672SP.

Domain	Sample global/national measures
Law enforcement, civil safety and security[27]	national crime index, oil/gas safety indicators
Epidemiology, healthcare, medicine[28,29,30]	rates of global/national infections, hospital deaths, disease prevalence, etc.
Human physiology	body mass index, metabolic equivalence
Environment[31]	rates of change in world climate, global effect on living organisms, pollution rates
Business, stock market	Dow Jones Industrials, leading and lagging business indicators, consumer price index[32]
Sociology[33], economics[34,35], development of nations	quality of life index, poverty, wealth, development, industrialization indicators

Table 7.1 Global/national measures in several domains

(b) This survey tells that many indicators do not exist as single variables. Instead, they appear as amalgams of several key variables from each source (say a country, hospital or economic sector). Next, these variables are usually weighted by importance and synthesized into *composite*[36] quantitative models[37,38]. This composite approach contrasts with the framework's intention here to aggregate one key variable at a time, say incident occurrence, across several airports.

[27] See, for example Munda, G. , Nardo, M., (2005). "Non-Compensatory Composite Indicators for Ranking Countries: A Defensible Setting", Institute for Protection and Security of the Citizen, European Commission, Report No. EUR 21833 EN, Luxembourg.
[28] See World Health Organization (WHO), (2005). "What are the advantages and limitations of different quality and safety tools for health care", Regional Office for Europe's Health Evidence Network.
[29] Michalos, A. C., Sharpe, A., Muhajarine, N. (2006). "An Approach to a Canadian Index of Wellbeing", Draft, Human Development Report Office, Canada.
[30] Jacobs, R., Smith, P., Goddard, M., (2004). "Measuring performance: An examination of composite performance indicators", The University of York, Centre for Health Economics, CHE Technical Paper Series 29, United Kingdom.
[31] See the 2005 Environmental Sustainability Index Report, Yale Center for Environmental Law and Policy, Yale University, Center for International Earth Science Information Network, Columbia University, Benchmarking National Environmental Stewardship, Appendix A, Methodology.
[32] Consumer Price Index Manual: Theory and Practice (2004). International Labour Organization.
[33] Booysen, F., (2002). "An Overview and Evaluation of Composite Indices of Development", Social Indicators Research, Vol. 59, pp. 115-151, Kluwer Academic Publishers, Netherlands.
[34] Sharpe, A., (2004). "Literature Review of Frameworks for Macro-indicators", Center for the Study of Living Standards, Ottawa, Canada.
[35] Saisana, M., (2004). "Knowledge Economy Indicators (KEI)", Development of Innovative and Reliable Indicator Systems, Joint Research Center, European Commission, CIS8-CT-2004-502529 KEI
[36] See Appendix A for definition of *composite*, in contrast to *aggregate* indicator.
[37] Jacobs, R. , Smith, P. , Goddard, M. , (2004). "Measuring Performance: An Examination of Composite Performance Indicators", Technical Series Paper 29, Center for Health Economics, University of York, UK.
[38] Nardo, M., et al., (2005). "Handbook on Constructing Composite Indicators: Methodology and User Guide", Statistics Working Paper, STD/DOC(2005)3, Organization for Economic Co-Operation and Development.

Given this observation, the challenge then becomes deciding how many indicators are needed to express surface safety sufficiently well while preserving analytical simplicity to make them readily acceptable. According to the survey, in general, a single synoptic indicator would be ill-suited to capture at once an entire state-of-affairs. Instead, a master, or "headline" measure complemented with collateral ones might be more effective. With this model in mind, we must entertain the possibility of adopting more than one safety indicator, despite our strive for minimal format.

7.2 Selection criterion

Survey aside, a selection criterion might prove to be helpful accessory, guiding us in our choice of measure to become a nationwide indicator. Toward devising this criterion, we first identify specific characteristics making a measure a good candidate. Several attributes come to mind. The product of much thought, eleven of them are itemized in Table 7.2. The plan is to use them collectively as guidance in screening and ranking potential measures before accepting them as indicators.

	Characteristic	**Qualifier**
1.	measurable and computable	determination is possible; data collection is feasible with no or few restrictions imposed on computation
2.	representative	reflective and meaningful of the measured target
3.	sensitive	reactive to changes in the variable being measured
4.	aggregatable	data from many sources can be combined into a single value representative of the whole
5.	unique	no two indicators can represent the same concept
6.	verifiable	subject to independent validation and verification
7.	viewable and clear	easy to display and understand
8.	flexible	modifiable and improvable with low effort
9.	accurate	subject to reasonable margins of error
10.	reliable	provides results consistently
11.	mature	established, recognizable in practice

Table 7.2 Selection criterion

7.3 Selection process

The selection strategy devised here is outlined in Figure 7.2 below. It requires formulating an overall goal first, such as "assess the safety of the nationwide airport system". This goal is then partitioned into a finer sub goal(s), say "rely on incident data to size such safety". That sub goal, in turn, leads to specific questions whose answers are to satisfy the sub goal.

Next, in the same cascading fashion, one searches for information necessary to answer the questions. In the search, several measures are conjectured to characterize that information. Only those that meet best the selection criterion (Section 7.2) are retained as final candidates for promotion to indicators.

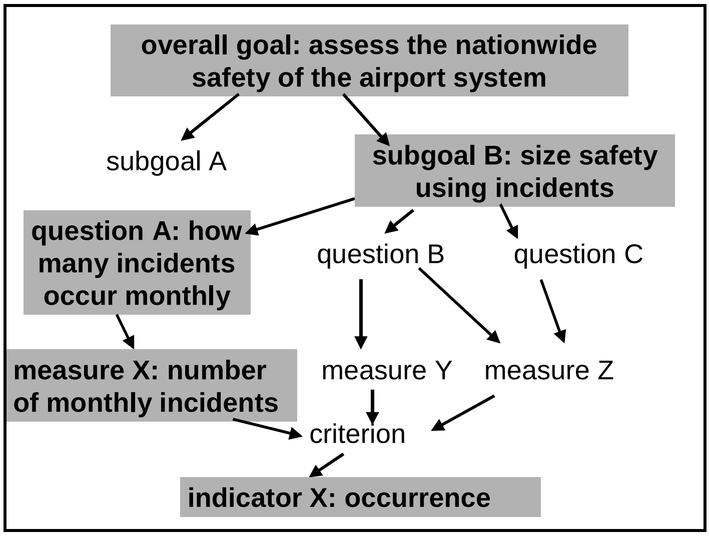


Figure 7.2 Selection process

Strategy-wise, to minimize the number of indicators needed, we anticipate the smallest number of questions a general audience might ask naturally when told about unsafe events. Framed that way, three priority questions come to mind. They are "how many, how often and how serious". According to the above selection process, candidate measures corresponding most directly to these questions are derived to be:

- occurrence, answering how many,
- recurrence, answering how often, and
- seriousness, answering how serious.

They are defined in the last row of Table 7.3 below. Finally, since they satisfy sufficiently well the Table 7.2 criterion, they are chosen for promotion to indicators.

goal	assess the safety of the nationwide airport system
sub goal	rely on incidents to size safety
questions related to sub goal	• how many incidents occur? • how often they recur? • how serious are they?
corresponding measures	• *occurrence*: incident count collected from each airport • *recurrence:* median days between incidents for each airport • *seriousness*: percentage of incidents considered serious

Table 7.3 Results of the selection process

7.4 Aggregation and display

Once measures are identified, they remain to be aggregated and displayed properly to become useful indicators. To reach this objective, much thought must go into characterizing them in practical form. The Occurrence, Recurrence and Seriousness subsections below offer several possibilities towards this realization.

7.4.1 Occurrence

Many options are possible in formulating the *occurrence* indicator. Counts, being simple in concept, are likely to stay meaningful in a variety of aggregation schemes. Some are described below.

7.4.1.1 Aggregate occurrence by local hour of the day

One could sum (aggregate) the count of all incidents by *local hour of day*, at a particular airport, regardless of incident type or severity. Consider Airports ABC and DEF, randomly selected from the data. Their resultant incident plots are shown in Figures 7.3 and 7.4, respectively. By themselves, ABC and DEF do not reveal any salient pattern, with numbers being so small, even over seven years. However, when the same process is cast at the national level a non-random bell-shaped pattern takes place as the time-distribution shown in Figure 7.5. This curve represents the *national model* of daily incident frequency. We see most of them occurring between 10 am and 3 pm local time, regardless of airport. This resultant pattern is quite different than ABC or DEF. Although the Figure 7.5 "aggregate nationwide airport" is virtual, existing only in the mind of the analyst, its pattern can serve as benchmark to assess how safety performance at a particular airport differs from the national norm over the course of the day.

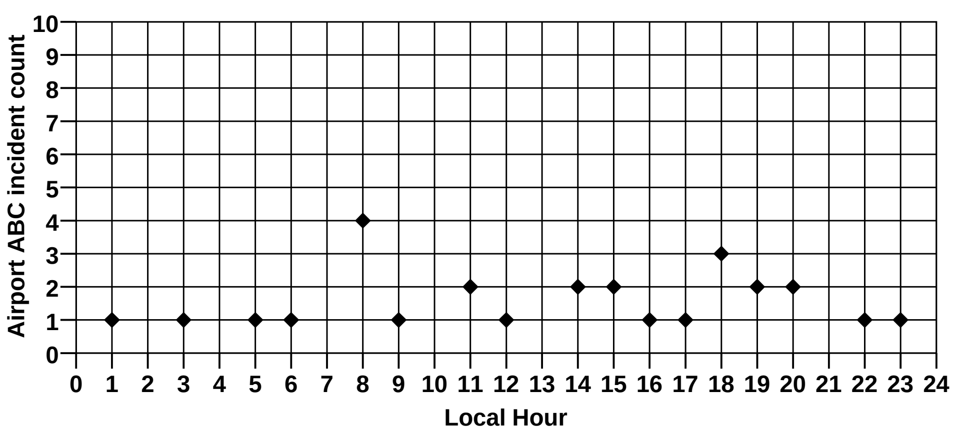


Figure 7.3 Airport ABC incidents aggregated by local hour of the day
(over 7 years from October 1999 to September 2006)

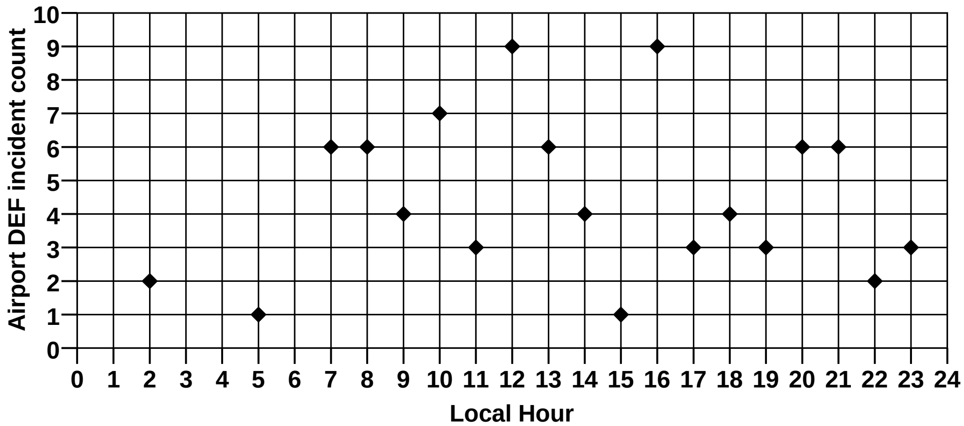


Figure 7.4 Airport DEF incidents aggregated by local hour of the day
(over 7 years from October 1999 to September 2006)

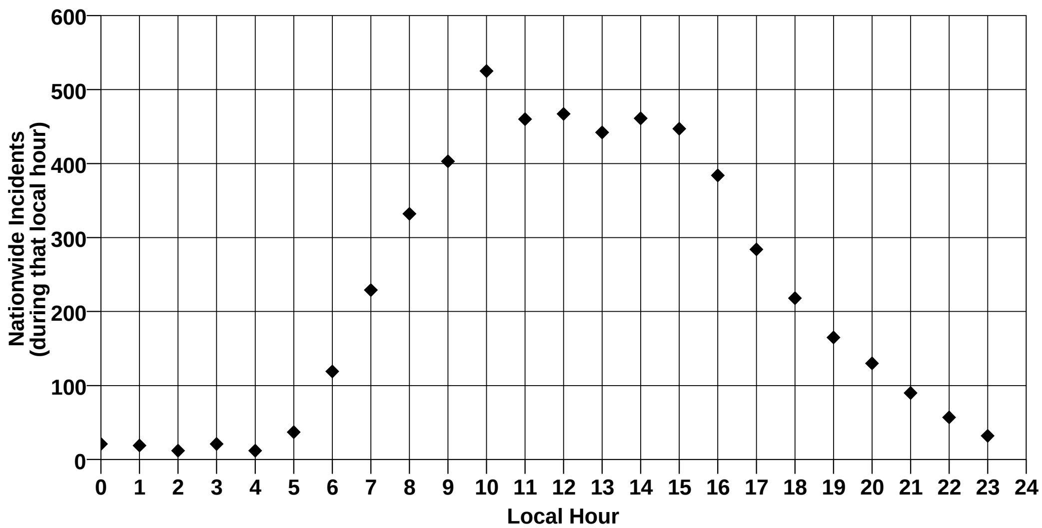


Figure 7.5 Nationwide incidents aggregated by local hour of the day
(Based on 5,367 incidents with valid occurrence times reported by 466 towered airports nationwide, over 7 years from October 1999 to September 2006.)

7.4.1.2 Aggregate occurrence by month of the year

Counts can also be aggregated by *month of the year*. The resultant seasonal trend is shown in Figure 7.6. January has a national low, then the counts climb in May-June with a dip in July. From there it declines back during the Fall and Winter. Traditional seasonal traffic activity is most likely the key factor affecting the shape of the curve. That is, higher seasonal traffic invites higher incident risk.

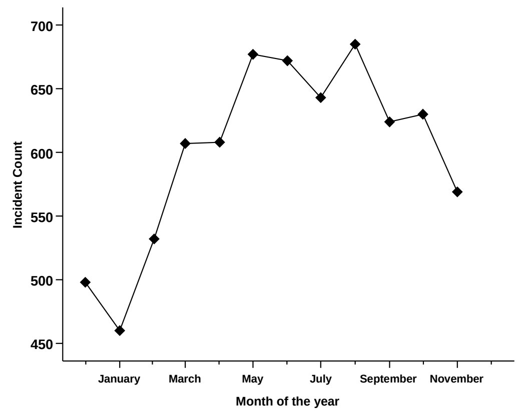


Figure 7.6 Nationwide incidents aggregated by month of the year

7.4.1.3 Occurrence control charts

So far, several displays have been suggested to bring out the data's salient features from different angles. Yet, despite any elegance they carry, these displays remain rudimentary, unable to estimate (or predict) incident trends over the years. To fill this gap, we devise a control chart showing not only how incident *occurrence* varies monthly over seven years, but also when it violates its own control safety limits. Traditionally used in industry to monitor the quality of goods and services, control charting is flexible enough to track safety performance as well, according to Brauer[39]. Its adaptation to incident counts is reasonable because the universal idea of monitoring quality matches the intuitive understanding of airport operations as a process delivering a quality level of safety to the users. Extending this reasoning, a control chart can therefore portray safety trends well enough with a wide multi-year, multi-month perspective, as depicted in Figure 7.7. Beyond this rationale, a control chart carries several distinct advantages, such as:

- easy to understand mainly because it is graphical;
- relies on measurement data, making it compatible with the empirical view adopted in our framework;
- supported by analytical rather than visual rules to decide when a process drops below expectation;
- widely used in quality and reliability circles in industry, and
- lends itself easily to further trending and forecasting analysis.

In Figure 7.7, the horizontal axis represents the seven years (84 months) while the vertical axis shows the total national incident count for each month, regardless of airport.

The control chart illustrating the occurrence indicator in Figure 7.7 is the "XmR" type, belonging to the simplest of formats. According to standard terminology, "X" means the "variable X" being measured (the monthly counts here) while "mR" stands for "moving Range" (the difference between current and preceding monthly count).

[39] Brauer, R. L., (2006). *Safety and Health for Engineers*, 2nd Edition, Wiley-Interscience, Hoboken, NJ, pp. 694-697.

The moving range is computed as an absolute (always positive) value. Figure 7.7 contains a central reference line representing the grand average number of incidents per month over seven years. In addition, an upper control limit is drawn based on 3 standard deviations (coined "3 sigmas") and the average of the "moving range R"[40]. The moving range chart is shown in Figure 7.8. Using its own trends, it serves to analyze more meaningfully the trends in the "X" chart (Figure 7.7). The moving range carries only a central line (an average) and an upper control limit.

The area inside the control limits in Figure 7.7 represents the leeway allowed for the "voice of the process". That is, this area conveys the limits of the natural fluctuations of a system operating as expected, under normal circumstances. Control limits are the critical feature making this chart more analytically useful than the preceding Figures 7.3 to 7.6. Limits are critical because they separate two types of variation in the data, which are:

- routine variation, also called "common cause", representing the ever-present random "noise" which reflects a certain number of incidents bound to occur in any active process (day-to-day airport operations, in this case) and

- exceptional variation, also called "special cause", representing a signal that the process is "out of control".

7.4.1.4 Special causes

Special causes can be isolated using one or more conditional tests. The rigor of these tests depends on how sensitive management decides to be when confronting airport incidents. For example, management concern might be raised if 3 sigmas are exceeded. Another concern might be present if 6 points in a row are trending up, meaning incidents are on the rise. If so, the following 9 cases in Figure 7.7 would deserve attention :

Particular month since Oct. 1999	Corresponding calendar month and year	Violations for Points
7	April 2000	Greater than +3 sigmas
8	May 2000	Greater than +3 sigmas
9	June 2000	Greater than +3 sigmas
10	July 2000	Greater than +3 sigmas
11	August 2000	Greater than +3 sigmas
12	September 2000	Greater than +3 sigmas
19	April 2001	Greater than +3 sigmas
20	May 2001	Greater than +3 sigmas
31	May 2002	Greater than +3 sigmas
31	May 2002	6 points in a row trending up

Table 7.4 Special causes for the *occurrence* indicator

[40] Wheeler, D. J., (1993). *Understanding Variation-The Key to Managing Chaos*, SPC Press Inc. , Knoxville, TN.

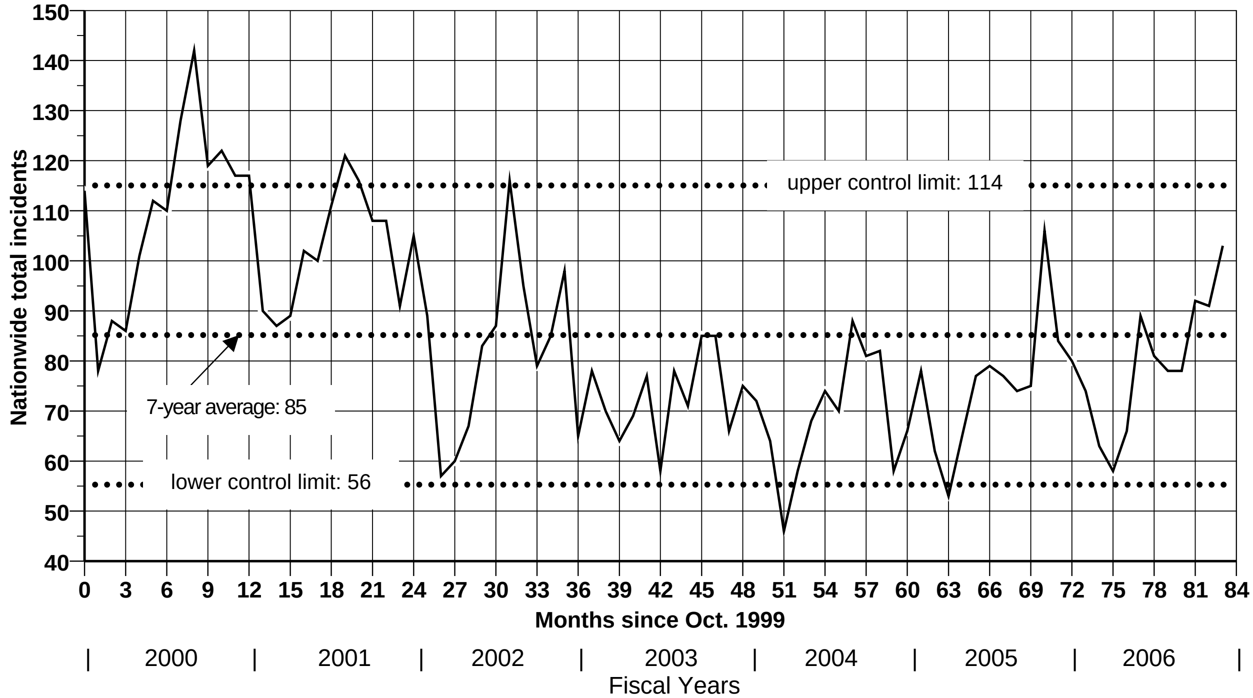


Figure 7.7 Incident occurrence as a nationwide aggregate safety indicator

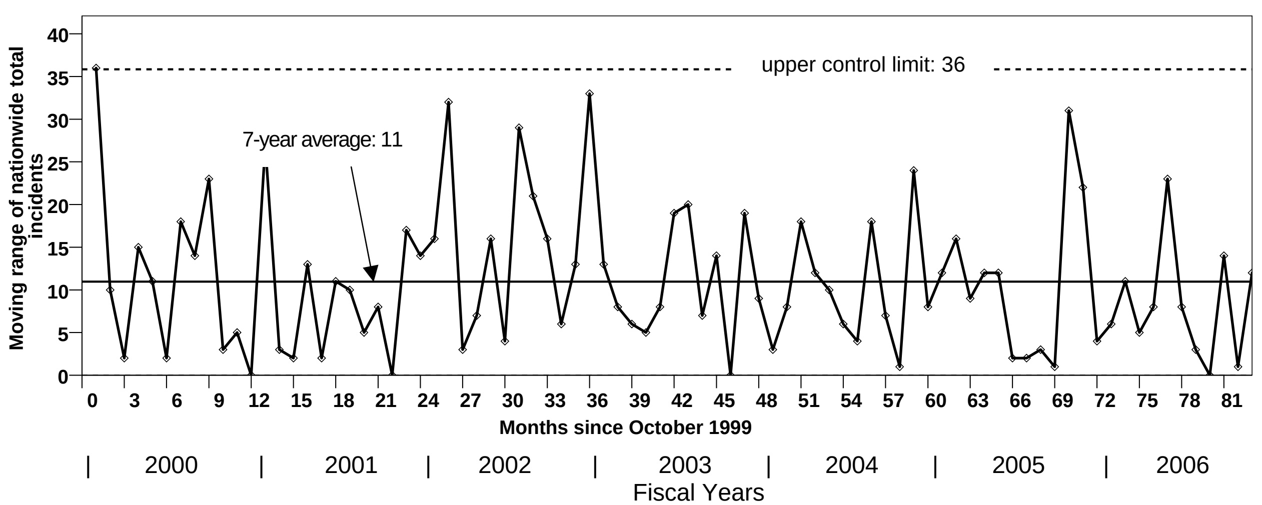
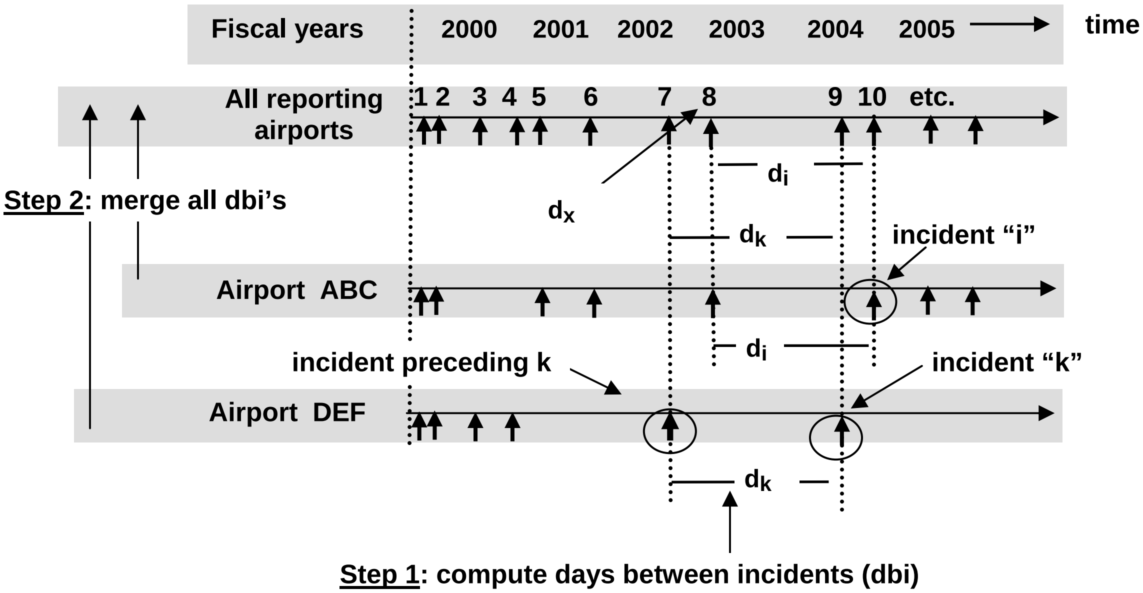


Figure 7.8 Moving range for incident occurrence

7.4.2 Recurrence

Constructing the *recurrence* indicator requires two steps. In Step 1, we compute the days between incidents (dbi) for individual incidents at each airport. In Step 2 we merge (aggregate) all the dbi's into a nationwide collection (labeled "All reporting airports"). Figure 7.9 illustrates the process for hypothetical Airports ABC and DEF.



.

Figure 7.9 The incident recurrence aggregation process

7.4.2.1 Aggregation

The term aggregation, as used here, differs from *pooling*. Pooling would require all incidents from all airports to be collected into a common "pool" first, then computing dbi's, regardless of airport source. For instance, 7,129 incidents occurred in seven years (2,555 days). Pooling them regardless of airport would result in an excessive number of about 7,129 / 2,555 = 2.79 incidents per day on average. Aggregation, on the other hand, pairs incidents strictly by airport source. This yields a different but less alarming value of about 47 days of median time between incidents at the typical airport in the United States. Although both computations are mathematically sound, the second choice is more realistic because it traces back each incident to its corresponding airport. Extending this claim, if an aggregate measure is to reflect a realistic view of safety it should not depend on an artificial "pooled" airport. Instead it should point back to individual actual airports as the real entities responsible for safe traffic. To enforce this condition, all dbi pairs in the aggregated collection remain identified by their airport origin.

Therefore, according to aggregation, incidents "8" and "7" in Figure 7.9 are not paired because each occurred at a different location (Airports ABC versus DEF). However, "9" and "8" are paired since both occurred at ABC. So the dbi's labeled d_i and d_k would be kept as genuine and d_x discarded as artificial.

7.4.2.2 Mean (average) and median times between incidents

Akin to the reliability term "Mean Time Between Failure" (MTBF)[41], recurrence represents the inverse of the safety failure rate (incidents per month). Like MTBF, the average time between incidents can be computed on an airport basis for each particular month in the seven years. By way of illustration, Table 7.5 lists the airports and corresponding number of incidents in a month randomly selected as "Month 26" (December 2001). In that month, 46 airports are found experiencing one or more incidents (see third column). The average and median of the dbi's for these incidents at each airport are shown in the last two columns.

Sample Month 26 : December 2001				
airport	**acronym**	**number of incidents**	**average days between incidents (dbi)**	**median days between incidents (dbi)**
1	DEF	2	11	11
2	ABC	1	81	81
3	PQR	1	414	414
4	.	1	95	95
5	.	1	25	25
6	.	1	350	350
7	.	1	185	185
8	.	1	111	111
9	JKL	3*	44**	15**
10	XYZ	1	blank	blank
11	.	1	300	300
12	.	1	185	185
13	.	1	8	8
14	.	1	65	65
15	.	1	195	195
16	.	1	39	39
17	.	1	120	120
18	PQR	3	13	6
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
43	.	2	185	185
44	.	1	125	125
45	.	1	202	202
46	.	1	36	36

* "3" comprises three incidents as follows: Incident 1: pilot deviation on 12/11/2001, Incident 2: pilot deviation on 12/11/2001, and Incident 3: pilot deviation on 12/26/2001.
** See Figure 7.11 for computation details.

Table 7.5 Sample monthly list of airports, incident counts, mean (average) and median days between incidents for December 2001

[41] Ayyub, B. M., McCuen, R. H. (2003). *Probability, Statistics, and Reliability for Engineers and Scientists* (2nd ed.), : Chapman & Hall/CRC, , Boca Raton, FL.

In the Table, the average and median columns are identical in cases where airports experience 1 or 2 incidents, as expected mathematically with rounding. According to the Table, such cases are the most frequent. However, some airports, like JKL and PQR, experience more than 2 incidents. For these cases, the average can be quite different from the median. For example, Airport JKL carries 3 incidents in that month, as shown in the Table. JKL's average is 44 days but its median is only 15, as computed according to the process illustrated in Figure 7.10 below.

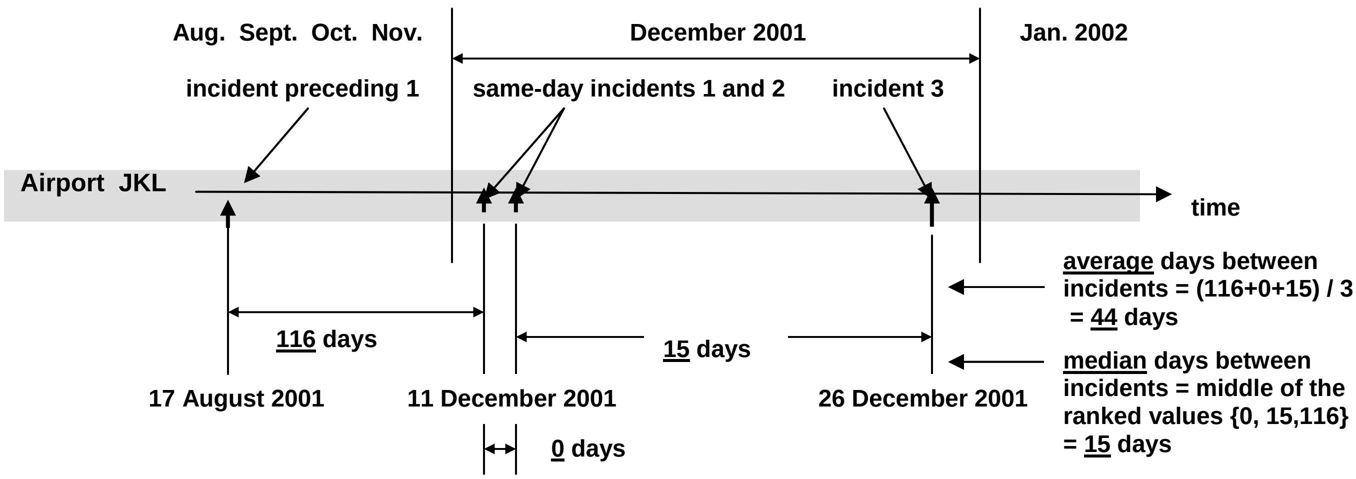
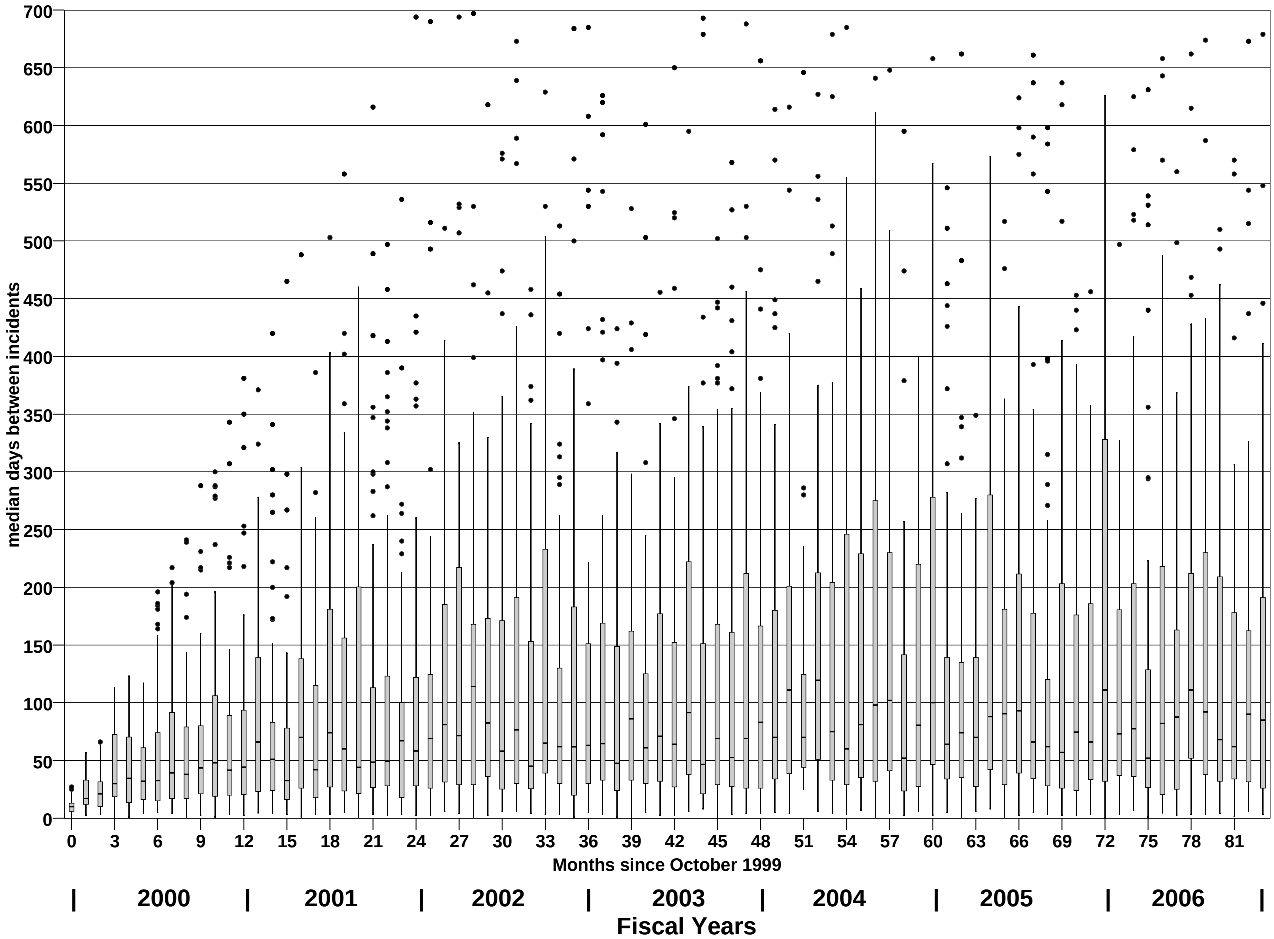


Figure 7.10 Computation of average days between incidents

Also note that Airport XYZ has one incident with blank average and median, given it is the first one found for that airport in the entire data. Finally, note that Airport ABC has one incident too, yet it carries an average and median of 81days. This is so because the average and median of one value (81) in that month both become the value itself. In that same case, the preceding incident (dated 22 September 2001) is used to compute the 81 days before the current incident (dated 12 December 2001).

The computation illustrated in Figure 7.10 can be summarized in a single display showing how airports' dbi medians are distributed within each of the 84 months over seven years. To do so, we use standard boxplots in Figure 7.11. This Figure shows a safety improvement because the median times grow generally longer with time, from month to month.



Legend:

Each bar represents several medians (one for each airport)

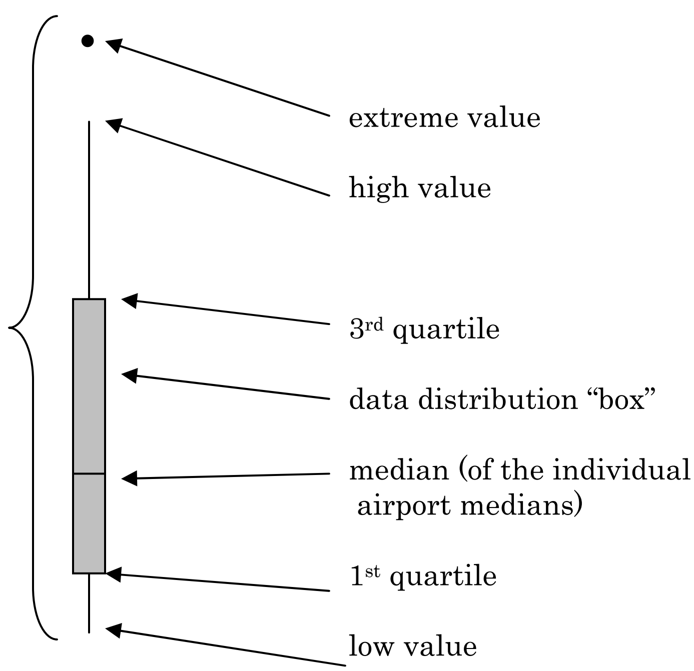


Figure 7.11 Boxplots for monthly distribution of average days between incidents

7.4.2.3 Recurrence control charts

In the same way as occurrence, the recurrence indicator is plotted as an XmR control chart in Figures 7.12 and 7.13, using monthly *median* days between incidents however instead of counts. The trend shows median safety is improving steadily over the last seven years, given the longer times shown between incidents. However, like in occurrence, special causes are present, as identified in Table 7.6 below.

Particular month since Oct. 1999	Corresponding calendar month and year	Violations for Points
0	October 1999	Less than -3 sigma
2	December 1999	Less than -3 sigma
78	April 2006	6 points in a row trending down
79	May 2006	6 points in a row trending down
80	June 2006	6 points in a row trending down
81	July 2006	6 points in a row trending down
82	August 2006	6 points in a row trending down
83	September 2006	6 points in a row trending down

Table 7.6 Special causes for the *recurrence* indicator

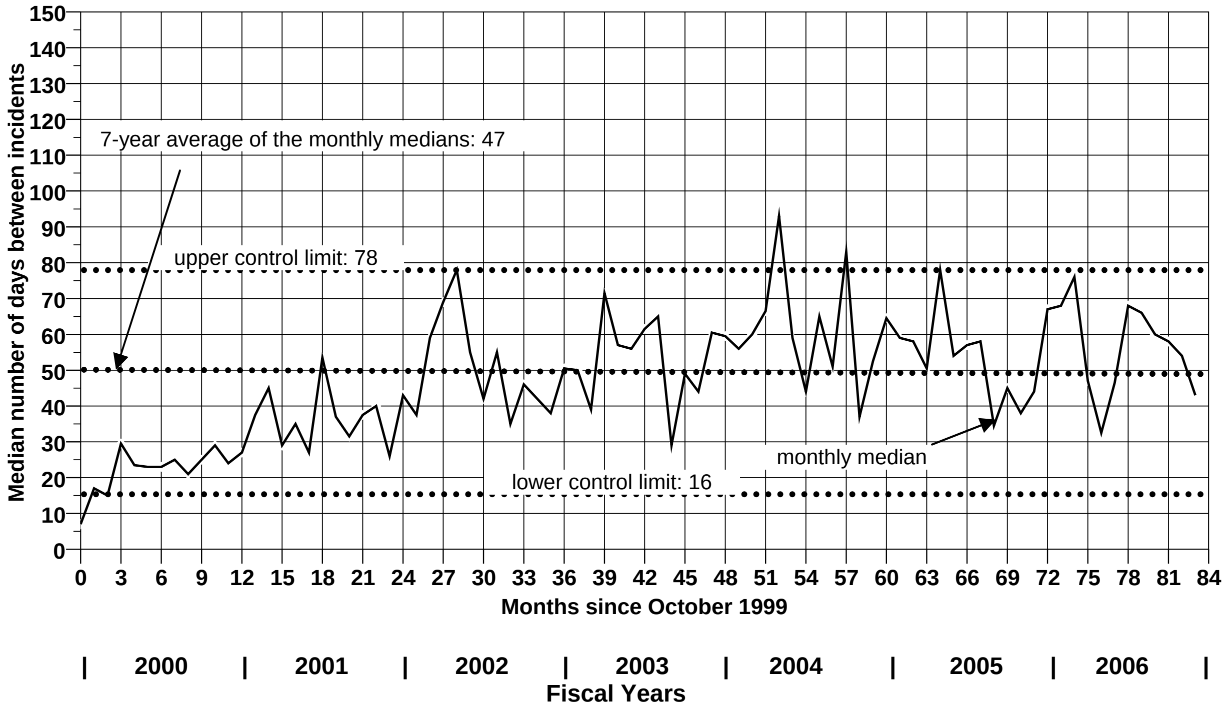


Figure 7.12 Incident recurrence as a nationwide aggregate safety indicator

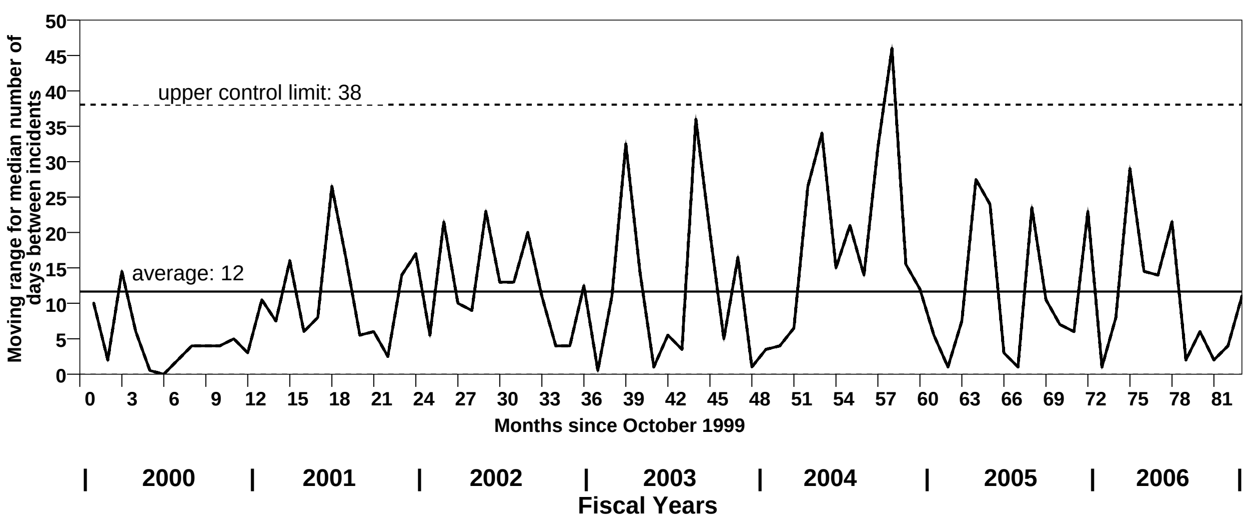


Figure 7.13 Moving range for incident recurrence

7.4.3 Seriousness

This indicator provides an important perspective to the methodology, since most of the incidents are not serious. Mathematically, it is kept simple by expressing it as a percent of total incidents for each month during seven years. Like the previous two indicators, it is plotted in Figure 7.14 and labeled "serious incursions". Except for one case in Table 7.7 below, the peaks never reach beyond 12 percent.

Particular month since Oct. 1999	Corresponding calendar month and year	Violations for Points
17	February 2001	Greater than +3 sigma

Table 7.7 Special causes for the *seriousness* indicator

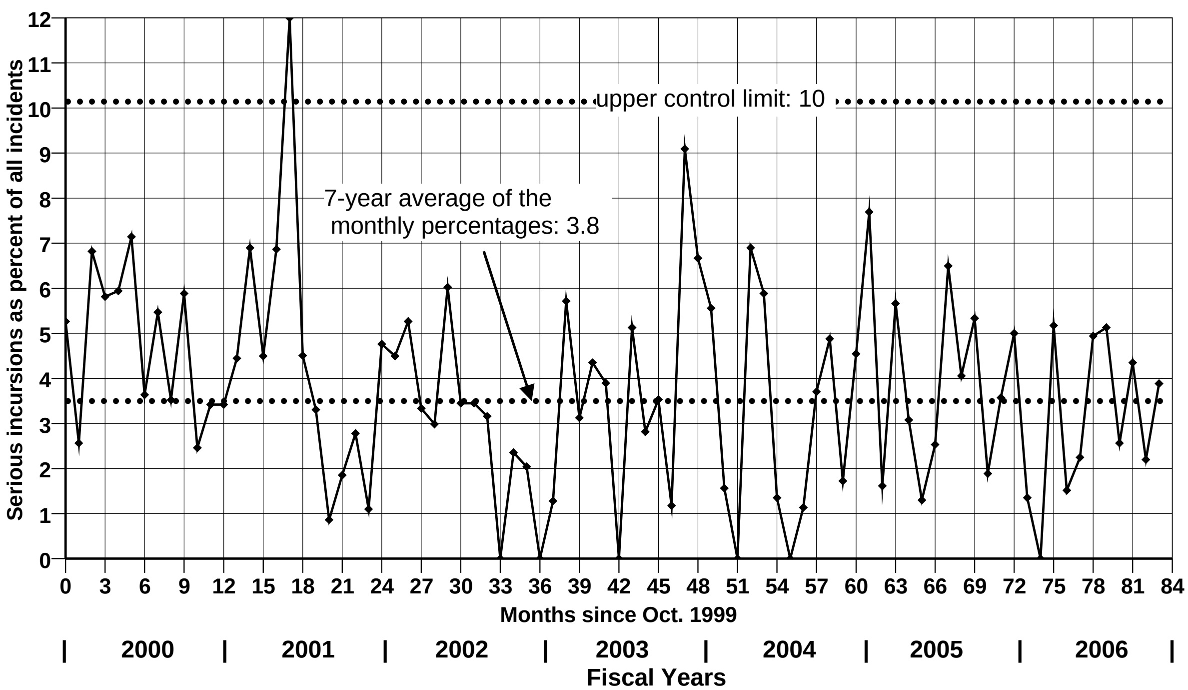


Figure 7.14 Incident seriousness as a nationwide aggregate safety indicator

7.5 Statistical summary

The 84 monthly summaries used in displaying the three indicators (see Figures 7.7, 7.12 and 7.14) are displayed as frequency distributions in Figure 7.15. The theoretical bell-shaped curve is sketched over the actual data bars to show how much the actual 84 values depart from the expected bell shape, (as predicted by the Central Limit Theorem[42] in statistics.) Also shown in 7.15 are the control limits bracketing the "voice of the process".

Finally, for the 84 summaries, Table 7.7 contains traditional "measures of location" (mean, median, etc.) as well as other, more modern versions like the 5% trimmed mean, for example. The "Definitions" column in the Table describes in words these statistical terms. Similarly, Table 7.8 provides "measures of dispersion" (variance, standard deviation, etc.)

---

[42] Freund, J. E., Perles, B. M, (2007). *Modern Elementary Statistics*, 12th Edition, Prentice Hall, NJ.

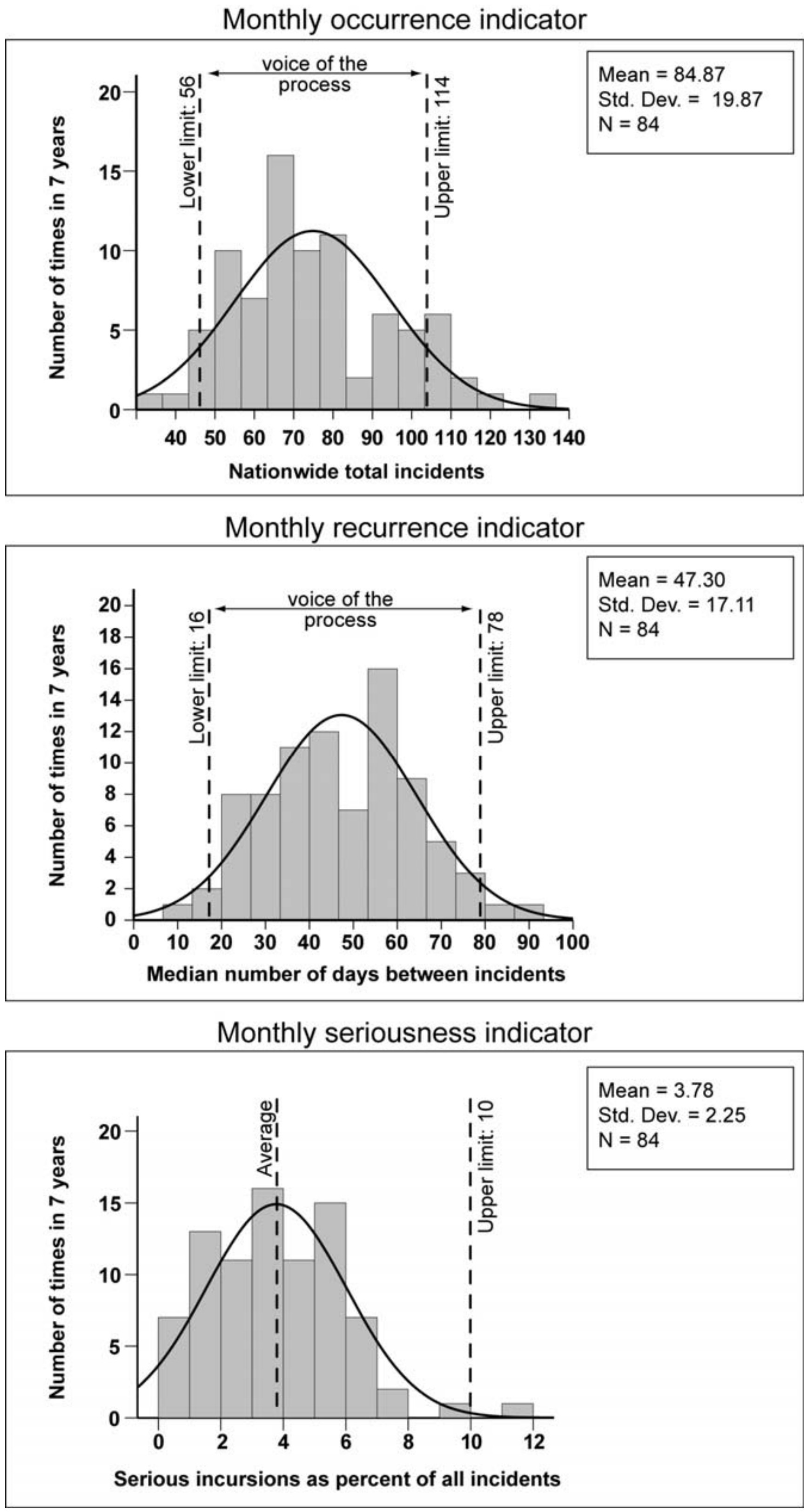


Figure 7.15 Frequency distribution of the three indicators

Descriptive statistics over 7 years	Definitions Units	Monthly Indicators occurrence number of incidents	recurrence median number of days between incidents	seriousness percentage of incidents
Mean (S.E)	arithmetic average of the 84 monthly summaries and (standard error of the average)	84.87 (2.168)	47.3036 (1.86707)	3.7777 (0.245)
95% Confidence Interval for the Mean	lower confidence limit	80.56	43.5900	3.2897
	upper confidence limit	89.18	51.0171	4.2657
5% Trimmed Mean	same as mean but computed with 5% of the lowest and highest ranked data values removed (trimmed)	84.21	47.1124	3.7040
Huber's M-Estimator	several resistant (robust) measures available in the Statistical Package for Social Sciences (SPSS). They are akin to the arithmetic average. However, they are more representative of the data[43], because each reflect a weighted mean less sensitive to outliers or skewness in the data. In general, values closer to the center of the data distribution are given more weight while distant values are given less weight.	82.25	47.2929	3.7016
Tukey's Biweight		81.65	47.0885	3.6504
Hampel's M-Estimator		82.83	47.1163	3.6915
Andrews' Wave		81.66	47.0820	3.6504
Median	the middle value in the data (50% percentile)	81.00	46.7500	3.5504

Table 7.8 Statistical results for the three indicators (measures of data location)

[43] Andrews, D.F., Bickel, P.J., Hampel, F.R., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, N.J.

		Monthly Indicators		
		occurrence	**recurrence**	**seriousness**
Descriptive statistics over 7 years	**Definitions Units**	**number of incidents**	**median number of days between incidents**	**percentage of incidents**
Variance	statistical measure of variation	394.766	292.819	5.057
Standard deviation	square root of the variance	19.869	17.11197	2.24877
Minimum	lowest value	46	7	.00
Maximum	highest value	142	93	12.00
Range	highest minus lowest value	96	86	12.00
Interquartile Range	same as the range but with the lowest and highest 25% of the ranked data values removed	30	24	3.18
Skewness[44]	measure of data distribution shift or “slide”	0.563	.081	0.564
Kurtosis[45]	measure of data distribution peakedness or flatness	-0.206	-.326	1.012

Table 7.9 Statistical results for the three indicators (measures of data dispersion)

[44] See Porkess, R. (2005). *Collins Internet-linked Dictionary of Statistics*, 2nd Edition.
[45] Ibid.

Appendix A: Definitions

This appendix defines key terms used in the previous sections. The definitions apply strictly within the airport safety context of this document.

A. Measurement-related terms[46]

a. *Count (Measure):* a numerical value representing the result of a measurement tally. Example: 15 incursions counted at Airport ABC by the end of Fiscal Year 2006. A count(s) states what is obtained from one or more soundings but leaves the audience to induce any trend beyond these numbers. During a measurement effort, counts are usually the simplest and easiest variables to report.

Example: data collected for Measure *M1* shows "1", "2", and again "1" runway incursions at a particular airport on August 2nd, September 7th, and September 8th, respectively, with no measurement frequency or trend stated explicitly beyond these counts.

b. *Rate:* a count divided by a factor deemed useful in adding perspective to the count.

Example: the phrase "15 incursions per 100,000 operations" becomes a *rate* of 0.0015 incidents per operation. Here, traffic load (operations) supplies the perspective.

c. *Metric:* Metrics belong further up the measurement pyramid. Unlike a count, a metric is a pointer or guide that stipulates a set of repeated, periodic measures *metered* in time.

Example: Extending the *M1* example, the metric *M2* consists of a chart showing an increasing number of incursions, going from 12 to 20, during the last three periodic readings. Because metrics are tracked repeatedly, they might point to patterns, correlations, and similar trends useful in safety assessments, predictions, and decisions.

d. *Indicator:* Indicators claim more sophistication than metrics. They represent metrics but with an upper and/or lower control limit(s) assigned to the tracking. The limits warn to take action on those data points falling outside the limits, as would be shown on a typical control chart.

Example: Indicator *M3* shows defects to be 9, 15, and 20, increasing beyond the indicated maximum prescribed limit of 10.

[46] Source: Press, J. (2003), A Measurement Framework for Air Traffic Safety Metrics, DOT FAA CT-TN04/10.

e. *Aggregate indicator:* one that combines a *same* variable from several geographical facilities (e.g. mean time between incidents at all airports) into a single figure metered daily, monthly, or yearly.

Example: a runway incursion count collected at each facility and aggregated (combined) into a nationwide value (say as a sum or average) to be recorded daily, monthly, etc.

f. *composite indicator:* In contrast to an aggregate indicator, a *composite* one would combines *more than one* variable (e.g. weather conditions, pilot proficiency, human factors, runway configuration, etc.) into a single numerical measure of safety.

B. Incident-related terms[47]

a. Incident: any airport safety event classified as a surface incident, runway incursion or collision. incidents are events that involve direct or potentially direct effects on the safety of aircraft operations and of persons involved in those operations. Accidents result in death or serious injury to a person in, upon, or about the aircraft, or in substantial damage to the aircraft itself. In contrast, incidents are less serious events. However, they are assumed to represent unfavorable safety events. As a result, they are traditional measures of the notion of safety, in the absence of air traffic control induced accidents which are very rare.

b. Surface incident: any surface event where unauthorized or unapproved movement occurs within the movement area or an occurrence in the movement area associated with the operation of an aircraft that affects or could affect the safety of flight. Surface incidents result from Pilot Deviations (PDs), Vehicle/Pedestrian Deviations (VPDs), or Operational Error/Deviations (OEs/ODs). A surface incident may occur on a runway or a taxiway.

c. Runway incursion: any occurrence on a runway at an airport involving an aircraft, vehicle, person or object on the ground that creates a collision hazard or results in a loss of separation with an aircraft taking off, intending to take off, landing, or intending to land. Runway incursions result from pilot deviations, operational errors, and vehicle or pedestrian deviations. A runway incursion has to have a collision hazard or a loss of separation. The FAA categorizes runway incursions in four categories depending on the potential for collision. These categories are:

A- Separation decreases and participants take extreme action to narrowly avoid a collision.
B- Separation decreases and there is a significant potential for collision.
C- Separation decreases but there is ample time and distance to avoid a potential collision.
D- Little or no chance of collision but meets the definition of a runway incursion.

[47] Source: FAA Order *7210.56C Air Traffic Quality Assurance,* August 15, 2002, prepared by the Air Traffic Evaluations and Investigations Staff, AAT-20.

These U.S. categories are to be superseded soon by the following International Civil Aviation Organization (ICAO)[48] set:

A- A serious incident in which a collision was narrowly avoided.
B- An incident in which separation decreases and there is a significant potential for collision, which may result in a time critical corrective/evasive response to avoid a collision.
C- An incident characterized by ample time and/or distance to avoid a collision.
D- Incident that meets the definition of runway incursion such as incorrect presence of a single vehicle/person/aircraft on the protected area of a surface designated for the landing and take-off of aircraft but with no immediate safety consequences.
E- Insufficient information inconclusive or conflicting evidence precludes severity assessment.

d. Collision: contact between two aircraft or between an aircraft and other vehicle or obstacle.

e. Pilot deviation: actions of a pilot that result in the violation of a Federal Aviation Regulation or a North American Aerospace Defense (Command Air Defense Identification Zone) tolerance.

f. Vehicle and/or pedestrian deviation: any entry or movement on the airport movement area by a vehicle operator or pedestrian that has not been authorized by air traffic control (includes surface incidents involving aircraft operated by non-pilots, like mechanics).

g. Operational error: an occurrence attributable to an element of the air traffic system in which:

(1) Less than the applicable separation minima results between two or more aircraft, or between an aircraft and terrain or obstacles (e.g., operations below minimum vectoring altitude (MVA); equipment / personnel on runways), as required by FAA Order 7110.65 or other national directive; or

(2) An aircraft lands or departs on a runway closed to aircraft operations after receiving air traffic authorization.

(3) An aircraft lands or departs on a runway closed to aircraft operations, at an uncontrolled airport and it was determined that a NOTAM regarding the runway closure was not issued to the pilot as required.

h. Operational deviation: an occurrence attributable to an element of the air traffic system in which applicable separation minima as referenced in paragraph 5-1-1a, Operational Error was maintained, but:

(1) Less than the applicable separation minima existed between an aircraft and adjacent airspace without prior approval; or

(2) An aircraft penetrated airspace that was delegated to another position of operation or another facility without prior coordination and approval; or

[48] See ICAO (2006). "Manual for Preventing Runway Incursions", Doc 9870, AN/463, International Civil Aviation Organization, Montreal, Canada.

(3) An aircraft penetrated airspace that was delegated to another position of operation or another facility at an altitude or route contrary to the altitude or route requested and approved in direct coordination or as specified in a letter of agreement (LOA), pre-coordination, or internal procedure; or

(4) An aircraft is either positioned and/or routed contrary to that which was coordinated individually or; as specified in a LOA/directive between positions of operation in either the same or a different facility; or

(5) An aircraft, vehicle, equipment, or personnel encroached upon a landing area that was delegated to another position of operation without prior coordination and approval.