Statistical Consulting Centre

Department of Mathematics and Statistics
The University of Melbourne
Victoria 3010

Telephone: (03) 8344 6995
Facsimile: (03) 8344 7499

Confidential report for

Sydney Airport Corporation Limited

Job No. E/2210

Report no. 815

17 May 2010

Methodological Review of the ACCC Airport Monitoring Report

Sue Finch PhD AStat
Marnie Collins BCom BSc (Hons) GStat
Ian Gordon PhD AStat

The Statistical Consulting Centre was established in 1984, to provide a high quality consulting service in statistics to business, industry, government and university researchers.

Our specializations include the fields of biostatistics, epidemiology, complex surveys, advanced data analysis, experimental design and statistical legal work. All our consultants have honours or higher degrees in statistics.

The statistical services the Centre provides include:

- assistance with designing a research project , whether an experiment or a survey;
- guidance in the interpretation of data;
- analysis of results;
- effective reporting and communication of research results;
- expert witness work on statistical matters;
- courses for business, industry, government and the University.

The Centre's aims include promoting the practice of appropriate and informative statistical methods, increasing the awareness of the importance of statistical expertise, and fostering mutually beneficial contacts between the University and the wider community.

Methodological Review of the ACCC Airport Monitoring Report

Sue Finch PhD, AStat
Marnie Collins BCom BSc(Hons) GStat
Ian Gordon PhD, AStat

17 May 2010

Executive Summary

- This report contains an evaluation of three surveys conducted as part of the ACCC's annual Airport Monitoring Report. The surveys are a passenger survey, conducted by individual airports, and an airline survey and a border agency survey, conducted by the ACCC. The surveys were evaluated using information publicly available through the ACCC website. The purpose of the surveys is to provide an evaluation of the quality of service provided by airports.
- The guidelines for carrying out the surveys and the annual report describing the surveys and findings do not contain sufficient information to allow a proper evaluation of the quality of the data provided by the surveys.
- The guidelines and the report do not clearly define the population of interest for any of the three surveys.
- We recommend that the design of the sample, the survey instrument, survey methodology and procedures, response to the survey, and sample sizes be clearly documented in the annual report to provide evidence of the quality of each of the surveys.
- The published guidelines for the passenger survey are open to interpretation in such a way that there could be important differences between the data collected at different airports.
- The summary measures for each of the surveys are based on averaging responses measured on an ordinal scale; this is not appropriate. We recommend the use of summary measures appropriate to ordinal scales.
- Many so-called "overall" measures combine information from responses to different questions, and also from different sources. These measures are not defined, and as such, are not interpretable.
- The combination of responses obtained from different surveys of different populations into an overall measure is inappropriate; it is not defined and is uninterpretable.
- The ACCC report aims to support comparisons between airports and over time. Such comparisons cannot be reliably made without statistical information to support statistical inference. We recommend reporting

standard errors or confidence intervals for estimates obtained from the surveys and for comparisons of interest.

Table of contents

1 Introduction

The Australian Competition and Consumer Commission (ACCC) monitors the quality of services at major airports in Australia. The monitoring involves data collection from five different sources: airport operators, AirServices Australia, passenger surveys, airline surveys, and border agency surveys. The results obtained from the different surveys and sources are "aggregated to give an overall view of the quality of service provided by airport operators" (ACCC Airport Monitoring Report 2008–09, p.59).

The results of this monitoring are presented in an annual report, available on the internet; this is called the ACCC Airport Monitoring Report. It is prepared by the ACCC on the basis of information from the five sources described above. The ACCC collects objective data, such as the number of seats in a terminal, and carries out the survey of airlines and surveys of border agencies. The major airports are required to conduct passenger surveys using guidelines provided by the ACCC.

Sydney Airport Corporation Limited sought advice from the Statistical Consulting Centre (SCC) on the monitoring methodology used by the ACCC. Sydney Airport has longstanding concerns about the robustness, rigour and reliability of the ACCC's methodological approach to monitoring the quality of service.

The Statistical Consulting Centre (SCC) was contracted to undertake a statistical review of the monitoring of quality of service carried out by the ACCC to assess the performance of Australia's five major airports, as reported in "Airport Monitoring Report 2008-09". The focus of the review is on the methodologies for the passenger, airline and border agency surveys; each is reviewed separately. This entailed an assessment of the methodology, including sampling, data analysis and reporting techniques.

In relation to the surveys, the SCC was provided with:

- The ACCC Guidelines (http://www.accc.gov.au/content/index.phtml/itemId/671400)
- The ACCC Airport Monitoring Report for 2008-2009 (http://www.accc.gov.au/content/index.phtml/itemId/917763)

All publicly available information about the methodology and monitoring results is provided on the ACCC's websites described above. As mentioned above the airport corporations carry out the passenger surveys. Sydney Airport, for example, contracts an independent consultant to carry out the

survey. This review is based on the publicly available information about the surveys; the SCC did not have access to other information that may have been provided by the ACCC to airports.

The abbreviations used in this report are as follows:

ACCC	Australian Competition and Consumer Commission
AC&BPS	Australian Customs and Border Protection Service
AQIS	Australian Quarantine and Inspection Service
DIAC	Department of Immigration and Citizenship
DTL	Domestic Terminal Lease
SCC	Statistical Consulting Centre

The structure of the report is as follows. Section 2 describes key matters to consider when evaluating the quality of a survey. Sections 3, 4 and 5 review each of the surveys respectively. Section 6 provides some comments about general issues arising from the review.

2 Key areas for evaluation

The purpose of each of the surveys is to collect data from a sample (such as a group of passengers at an airport) that can be used to make inferences about a population (such as all passengers at an airport in a given time frame). The accuracy and reliability of the information provided by the sample about the population depends on many different aspects of the way the survey is designed and carried out. In this section, we set out the main points to consider when evaluating the quality of a survey. They are:

1. What is the population of interest? There is typically a population about which an inference is sought. The elements of the population could be people, households, companies, or some other defined "unit".
2. What was the "sample frame" for the survey? The sample frame is the set of units from which the sample is actually drawn. The "sample" is made up of a set of units from the population, from which we have data relevant to the inference sought.
3. How was the sample selected? That is, what process was followed to obtain the sample?
4. How was the survey administered? Was it a mail survey, or phone, or face-to-face, or was some other means used?
5. What was the response rate? What is the potential for response bias?
6. Is there any information on non-responders (units chosen to be in the sample, but from whom no survey data are available)? If so, what does the information suggest about response bias?
7. How were potential participants approached? Was there a covering letter or other documentation provided with the survey? Is its wording fair and does it seek an honest and frank response from those surveyed?
8. Is the questionnaire itself well designed? Are the layout and wording of questions conducive to clear and unambiguous responses? Do the questions allow for the full range of possible facts and opinions?
9. What evidence is there about the quality of the responses, from inspection of the survey forms returned?
10. Are the results of the survey reported clearly? Are estimates of population quantities reported with an indication of the precision of the estimate, for example, a "confidence interval"?
11. Taking all of the above considerations into account, what weight should be attached to the reported results of the survey? How reliable are the results?

Specific answers in relation to these matters can lead to indications of good or poor survey quality. Information about the particular way in which a survey is designed, conducted and analysed is essential to understanding the extent to which the results can be relied on; without such information it cannot be assumed that the survey has been conducted in an appropriate way.

In subsequent sections of this report, we comment on each of the three surveys with respect to the matters to consider outlined above. We focus first on the passenger survey and provide a detailed discussion of the relevance of each of the matters to a good quality survey. This more general discussion is not repeated for the other surveys.

3 Passenger survey

Surveys of passengers are the responsibility of the airport corporations.

> "The ACCC takes the view that such surveys can be undertaken 'in-house' by airport operators provided that the operators consult closely with the ACCC on both the contents of the survey and the methodology used. From the ACCC's perspective the surveys need to gather information that is relevant to the ACCC's quality of service monitoring program. The ACCC needs to be satisfied that the data collected, and the methodology and processes used, target priority areas and are statistically robust."
>
> (Airport quality of service monitoring guideline, ACCC, October 2008, p. 6-7)

The guidelines provided by the ACCC focus on the questions to be asked of passengers; there is scant information about other aspects of the survey methodology.

The description of the passenger survey provided in the Airport Monitoring Report is:

> "The passenger perception surveys arranged by each airport differ in their coverage and detail; however, these surveys must provide information consistent with that specified in the regulations and quality of service guideline. The areas covered include passenger check-in, security clearance, government inspection, gate lounges, washrooms, baggage processing and trolleys, signage and way finding, car parking and airport access for arriving and departing passengers."
>
> (ACCC Airport Monitoring Report 2008–09, p.14)

Other references to the passenger survey in the report are those describing the results.

The ACCC's website[1] provides an Excel worksheet that is a template for recording survey results. This also gives some information about the ACCC's expectations about the survey methodology:

> "The ACCC needs to be satisfied that the data collected, and the methodology and processes used, target priority areas and are statistically robust. The ACCC therefore expects that a full description of survey methodology is provided as a complement to the collated results including:
> - sample size

[1] http://www.accc.gov.au/content/index.phtml/itemId/671400

- mix of international/domestic passengers surveyed
- mix of arriving/departing passengers surveyed
- who conducted the surveys (eg the airport operator or otherwise)
- details of the scoring system used (eg 1=very poor, 5=excellent)"

(Airline survey quality of service monitoring templates for 2008-09.xls)

With this background information about the passenger survey, we consider each of the key areas for survey evaluation described in Section 2.

3.1 The population of interest

There is typically a population about which an inference is sought. The elements of the population could be people, households, companies, or some other defined "unit". While it is not specified explicitly, it is implied that the population about which inference is sought from this survey is airport passengers. For example, results are reported for passengers overall and also separately for domestic terminal passengers, international terminal passengers, and other airport services passengers. The definition of "other airport services" is not explicitly provided; Michael Samaras from Sydney Airport advised the SCC that it includes regional services.

The description of the population should be more precise. The ACCC provides an annual report implying that the population of interest is all passengers using a particular airport in a given year. The definition of the population is relevant to how the survey is designed and carried out. If the survey is carried out in a single week, for example, we need to ask - are inferences needed for all potential passengers passing through the airport during the time that the survey was conducted or are the inferences needed for a broader time period? It cannot be assumed that a survey conducted in a short time frame will appropriately sample the population of all passengers in a given year.

The definition of a passenger is not straightforward. A passenger could be defined as an individual; a passenger may make one or more trips through a particular airport in a given year but each individual only appears in the population once. Alternatively, a passenger could refer to an individual making a particular trip through a particular airport; an individual person has the potential to be part of the population more than once. This means that most people who fly at a given airport are included at least twice in the population of passengers (arriving and departing in the same year). However, a passenger could be defined as an individual making a round trip

at a particular airport; in this definition people could appear once or more in the population of passengers. This distinction has implications for the survey design and methodology. For example, passengers are asked questions about departure experiences, questions about arrival experiences, and questions that could apply to either arrivals or departures. However it is not clear if all questions are relevant to passengers surveyed at a particular time; this depends on the definition of a passenger. Implicitly, the survey appears to use the second or third definition of a passenger. However this is only conjecture. Additionally passengers might only refer to adults or could include children above a certain age. Again this is undefined.

The survey guidelines should provide an explicit definition of a passenger that is used by all airports in conducting the passenger survey. The report of the survey should also provide this definition.

3.2 Sample frame

The sample frame is the collection of passengers that could be sampled, from which the actual sample may be taken. Without an explicit and careful description of the population of interest, we cannot define an appropriate sampling frame. If we wish to make inferences about all passengers passing through a given airport in a particular year, then the sampling frame includes all those individuals.

As described above, the ACCC Airport Monitoring Report states that "the passenger perception surveys arranged by each airport differ in their coverage and detail" (p.14). However it is not explained what is meant by coverage and detail. Potentially it means that the surveys may have different coverage of passengers – and hence may be sampling different populations. They could, for example, be using different definitions of a passenger. The ACCC highlights the importance of assessing the various airport surveys in terms of their methodologies, and requires a full description of the methodology from the airport corporations. This description, including the sampling frame used at each airport, should be part of the Airport Monitoring Report.

There are a number of possible sampling frames that could be used for a survey of passengers; for the moment we will assume that the population of passengers refers to all people either arriving at or departing from a given airport on a particular trip in a given year. As described above, this means that most individuals who fly at a given airport are included at least twice (arriving and departing in the same year). In principle, airlines could

provide lists of all departing and arriving passengers (ignoring issues of confidentiality for the moment); such a list would provide an appropriate sampling frame.

An obvious method for potentially accessing the population of interest would be to intercept passengers at airports; this could be based on locations through which all passengers must pass such as arrival and departure lounges. The practicalities of such an approach are not known, and although this provides an appropriate frame for sampling passenger trips, the implementation of an appropriate sampling method using this frame may not be straightforward.

There are other less appropriate sampling frames that might have been used for the survey. Telephone surveys are a traditional way of accessing a reasonable proportion of the Australian population; the sampling frame would have a number of biases in terms of accessing passengers but might be an inexpensive method of conducting a survey. A household survey is an alternative way of sampling the Australian population. These alternatives are not recommended, but without explicit details about the survey methodology, they cannot be ruled out.

Ideally similar sampling frames would be used at each airport. Appropriate sampling frames should be described in the ACCC guidelines, and the actual sampling frames used must be reported with the survey results.

3.3 Sample selection and sample size

The quality of a survey rests heavily on the method by which the sample is selected. Characteristics of a good sampling method are not always well understood. In many contexts, an appropriate sampling method can be practically difficult to implement and hence great care is needed in designing and carrying out the sampling plan. Even with a well-executed suitable sampling plan, the precision of the survey results depends on the sample size.

The ACCC guidelines do not provide any instructions to airports about sampling procedures or sample sizes for the passenger survey. There is no information regarding these issues in the most recent Airport Monitoring Report.

Sample selection should be based on a method of probability-based sampling; this means that all members of the population should have a known and non-zero probability of being selected. A simple random sample is an example of

a probability-based sample; a simple random sample of passengers could be obtained if a list of all passengers was available and a subset of n passengers from the list was selected in such a way that all possible subsets of size n were equally likely. There are other methods of probability sampling that do not require equal probabilities. Probability-based sampling methods are the basis for using statistical theory to draw inferences about populations from samples.

In many practical survey situations, non-probability sampling methods are used. Non-probability sampling methods include convenience sampling where individuals are sampled in an ad-hoc way and quota sampling where individuals are selected to fit some predetermined criteria. Sometimes the claim is made that a "representative" sample has been obtained; this is often a vague claim and may be based on the view that personal biases have been avoided in selecting people to survey. Although survey staff may attempt to avoid bias, there are many ways in which bias might arise, and some of these may be quite subtle and unintentional. The term "random sampling" is also sometimes inaccurately used to refer to a non-probability sample selected without conscious bias.

Non-probability sampling methods, although widely used, have many drawbacks; the estimates of population characteristics based on non-probability samples cannot be considered to be objective, and their precision is unknown.

There is potential for quite different sampling methods to be used in different airports. It is not essential that the same method be used in all airports. However, using the same probability-sampling method at all airports would provide a sound basis for making comparisons between airports.

A common misconception is that the quality of a survey depends only on the size of the sample; large samples are regarded as implicitly high quality irrespective of the sampling methods used. If poor sampling methods are used, the size of the sample is irrelevant; the information obtained from the sample about the population is likely to be misleading and inaccurate.

A large sample size, in itself, does not guarantee reliability. The largest survey ever conducted, with a sample size of 2.4 million, made an estimate of a population percentage that turned out to be alarmingly different from the

true population figure: the estimate was 43%, and the population figure was 62%.[2]

In a well-designed and implemented survey, sample size becomes important in determining the precision of the survey results. The sample will provide an estimate of a population characteristic, for example, the proportion of passengers who regard the check-in waiting time as very poor. The sample proportion is the best estimate of the true population proportion. However, results will vary from sample to sample. A confidence interval provides a range of values for the population parameter that is consistent with the data from our sample. Hence it helps to indicate the precision in the results. Sometimes the precision of a survey is described in terms of a margin of error – this is half the width of the confidence interval.

The choice of sample size should be made by considering the desired level of precision in the survey results. As will be discussed further below, there is no information about the precision of the survey results in the Airport Monitoring Report. It is possible, for example, that the results from the different airports have quite different levels of precision. Without this information, any inferences about differences between airports or changes over time will depend on subjective judgement.

3.4 Survey administration

The ACCC does not provide any guidelines or instructions to airports about methods for the administration of the passenger survey; there is no description of the methods used in the Airport Monitoring Report. It is important to know how the survey was administered – was it a mail survey, phone, face-to-face, was an interviewer used or was it self-administered? These issues have a bearing on the survey quality is many ways. If, for example, the survey is self-administered we need to know what procedures are followed if English is not the language of choice for the individual sampled. There is more potential for missing responses or for misinterpretation of responses if the survey is self-administered than if an interviewer is involved. The method of administration can also have a bearing on the response rate, as discussed below.

[2] This is the famous Literary Digest poll of voting intentions in the 1936 US presidential election.

3.5 Response rates and response bias

The response rate to a survey indicates the proportion of the units sampled that participated in the survey. Assume, for example, the passenger survey sampled every *n*th passenger as they entered into a departure gate accessible only to passengers, and that passengers are approached by survey staff for interview. Non-response to the survey could arise, for example, from refusal to participate, language difficulties, failure to complete the survey instrument and failure of survey staff to approach a sampled passenger. The response rate should account for all sources of non-response, not simply refusals to participate.

More often than not, response rates to surveys are inadequate. It is difficult to be definitive about what constitutes an acceptable response rate under any circumstances. However, in almost all situations, a response rate of over 90%, for example, is likely to be sufficient to render the results reliable for quantitative conclusions.

Equally, it is difficult to say what constitutes an unacceptably low response rate under any circumstances. However, statisticians generally agree that there is a threshold below which response rates are unacceptably low. They may disagree about the actual value of the threshold.

Many writers decline to construct specific guidelines. However, Babbie writes: "A response rate of at least 50 percent is generally considered adequate for analysis and reporting. A response rate of at least 60 percent is considered good, and a response rate of 70 percent or more is very good".[3] In commenting on these sorts of guidelines, Lohr writes: "I believe that giving such absolute guidelines for acceptable response rates is dangerous and has led many survey investigators to unfounded complacency; many examples exist of surveys with a 70% response rate whose results are flawed."[4]

The SCC gives the general advice that a response rate of lower than 30% means that the results of a survey must be regarded as unreliable. The main concern is about possible response bias in the surveys considered here. A high response rate is vital: the higher the response rate, the less the scope for response bias. Conversely, in a survey with a low response rate, the scope for response bias is large.

[3] Babbie E.R. (1990). Survey Research Methods (2nd edition). Belmont, California: Wadsworth, page 182.

[4] SL Lohr. (1999). Sampling: Design and Analysis. Duxbury, California, page 281.

In a survey of passengers there are many potential sources of bias; the possibility of losing non-English speaking participants has already been mentioned. However a likely major source of response bias is refusal to participate. The way in which this is likely to arise will depend on the method of survey administration. Non-response may be higher in a self-administered survey than a survey where an interviewer asks the questions on the spot.

It cannot be assumed that the responses of individuals who choose to participate in a survey will be similar to those who choose not to participate. Passengers who refuse to participate may be more pressed for time than those who agree to participate; this could relate to their experiences of airport services. Passengers who refuse to participate may have more demands on them than those who agree; again the experiences of these different passengers may be quite different. Conversely, people who have had an unpleasant experience may be more inclined to spend the time responding than others.

There is no mention of response rates in the Airport Monitoring Report or what is considered acceptable in the ACCC guidelines. Response rate is not mentioned in the ACCC's reporting requirements for the passenger survey. In the report of the survey, there is no information about non-responders. It is possible that some information exists if the sampling frame is airline passenger lists. It is essential that the response rates and the various sources of non-response be reported in the annual report; without this information the quality of the information provided by any of the airports cannot be assessed.

3.6 Approach to potential participants

With the lack of detail about the survey methodology, the approach to sampled participants is not described. In order to avoid bias in the sample, all sampled participants should be encouraged as strongly as is reasonably possible to take part in the survey. The information provided to passengers at the point of recruitment should clearly describe the purpose of the survey and the expectations of the passenger. Honest and frank responses should be encouraged. With passenger surveys at different airports being conducted by different companies, there may be substantial differences in how passengers are encouraged to take part. It would be appropriate to provide this kind of detail in the guidelines.

3.7 Questionnaire design

The ACCC guidelines and report focus on the scope of issues to be covered in the passenger survey and broadly the type of questions to be asked. The passenger survey criteria for monitoring are detailed in Attachment A of the guideline document, and the measures are also listed in Appendix A2.1.2 of the Airport Monitoring Report. The Excel file on the ACCC website is the method by which the passenger survey data is provided to the ACCC by the airport operators. Again these indicate the measures of interest. None of these sources provide any information about the exact wording of questions in the passenger surveys. Again, there is no information regarding survey format or design. We do not know if the questions used by different airports are likely to measure the same underlying construct. We do not know, for example, if passengers were asked to give a response relating to a trip they were undertaking or to a trip just completed, or to evaluate all the trips they had taken through that airport during the year of interest. There is no information available on which to judge the quality of survey responses.

The Excel spreadsheets for providing data include a "not applicable" option; it should be clear, for example, if this was an option provided to passengers explicitly. As mentioned earlier, it appears that not all survey items are necessarily relevant to all passengers. For example, international passengers may have no knowledge of airport car parking facilities.

Again, an approach that ensures greater consistency across airports in design of the passenger survey is important; the guidelines as they currently stand allow for substantial variation in the questions asked of passengers.

3.8 Reporting of the results

The measurement scales used in the survey are described:

> "Surveys at most airports ask respondents to rate their level of satisfaction with the facilities on a scale from 1 to 5 (table 1.4.1).

Table 1.4.1: Ratings of satisfaction for airport facilities and services

1	2	3	4	5
Very poor	Poor	Satisfactory	Good	Excellent

(ACCC Airport Monitoring Report 2008-2009, p.14)

These measurement scales are ordered categories. The assignment of values 1 to 5 to this ordinal scale is entirely arbitrary. The ACCC indicates how the results for these measurements should be summarised:

> "The average ratings for each indicator in the passenger perception surveys are shown for each airport. The average ratings for domestic terminals and international terminals are presented over time where possible."
>
> (ACCC Airport Monitoring Report 2008-2009, p.14)

The average "scores" are treated as if they are continuous measures reflecting satisfaction.

This method of assigning values to the categories that participants use to respond to the questions is *arbitrary*. If an alternative method of assigning numbers to the categories was used, it could alter the observed averages. For example, consider assigning the following numeric values:

0	Very poor
5	Poor
15	Satisfactory
25	Good
35	Excellent

This is an example where differential spacing of the numerical values has been used. This could be a more reasonable assignment of scores, if it was perceived that a response of Poor was closer to a response of Very poor than to Satisfactory. (However the main point here is to illustrate the arbitrary nature of the assignment of numerical values to categories.).

Although the assignment of numerical values to categories is quite a common practice, it is inappropriate, without very strong assumptions about the nature of the ordinal scale. The measurements of satisfaction are best considered as ordered categories and hence using measures such as means to summarise the results is not meaningful. The essential point is that ordered categories remain categories, and it is impossible to reduce the distribution of a categorical variable to a single summary figure, without simplification. There are a variety of alternate and valid ways of interpreting the measurement scales.

Currently, the results of the passenger survey are reported as means without any indication of variability or error. When results are reported without any indication of variability (e.g. confidence intervals or standard errors), a reader might conclude that airports or time periods are substantially different when, in fact, the observable differences may be no more than would be expected by chance, or the differences may not have been precisely estimated. Estimated values of population characteristics should always be accompanied by a measure of precision. Statements about differences should be accompanied

by a summary measure of the difference as well as by the results of some method of statistical inference.

The ACCC states that the:

> "… quality monitoring program aims to gather and report data on criteria that facilitate assessments on changes in service quality over time as well as possible comparisons across airports."
>
> (ACCC guidelines, p.3)

No formal statistical inferences are reported for the survey; a minimum standard of reporting practice would include some form of statistical inference given the stated aims of the quality monitoring program.

The ACCC reports an overall rating of the quality of passenger services; there is no information about how this overall rating is obtained from individual survey questions. There are many different possibilities. Appropriate methods of obtaining an overall index must consider the ordinal nature of the measurement scales used for the individual items. In any case, an overall rating cannot be interpreted in any meaningful way if the way it is obtained is not explained. It may be more useful to focus on individual aspects of quality of service rather than an overall score; this would help to identify specific areas to target for improvement.

In this context, it is also noted that an overall rating is also reported that combines information from all five sources mentioned in Section 1 of this report. The derivation of this overall measure is not described; again, we cannot meaningfully interpret such a measure without an explanation of its construction.

The results of the passenger survey are presented graphically using line graphs and bar charts. The line graphs, for example, show a separate line for each airport and plot the average rating using labels of Very poor, Poor, Satisfactory, Good and Excellent equally spaced on the vertical axis. An example of this style of graph can be seen in Chart 2.2.7 on page 33 of the Airport monitoring report 2008-2009. This is a somewhat ambiguous representation, reflecting the underlying inconsistency that arises from averaging a categorical scale.

3.9 Overall quality of the survey

The ACCC makes some strong statements about the importance of robust survey results:

> "The ACCC needs to be satisfied that the data collected, and the methodology and processes used, target priority areas and are statistically robust. In this context, auditing and verification procedures will be important.
>
> The ACCC expects airport operators to cooperate in providing meaningful data for the quality of service monitoring program and to this extent it may require, under s.156 of the Airports Act, that the collated results of surveys be certified by statutory declaration. The ACCC also expects that a full description of survey methodology and raw data is provided as a complement to the collated results."
>
> (Airport quality of service monitoring guideline, ACCC, October 2008, p. 6-7)

In this context, the lack of detail provided about the survey methodology in both the guidelines and the report of the results is surprising. Without such information, the overall quality of the survey results cannot be judged to be adequate. Readers of the ACCC report should be provided with sufficient detail to be confident that the results are of sufficient quality.

There are many aspects of the design, methodology and analysis that are unspecified or unclear. Given that different airports conduct the different surveys, there is potential for there to be important differences between the surveys that could influence results. Ideally, a common methodology should be employed at each of the airports to ensure valid and fair comparisons between airports.

The matters that are referred to in the guidelines and survey results that are relevant to survey quality are not comprehensive. The implied description of the survey methodology is not a full description. Many of the matters discussed in this document are not included.

The survey results do not have a clear interpretation. Even if a valid, robust and consistent methodology was used across airports, the results as presented are uninformative for several reasons: a continuous numeric scale is imposed on a set of ordered categories; survey items are combined in an unspecified way to produce an overall measure of passenger assessed quality of service,

and there are no measures of precision to aid the interpretation and importance of observed differences between airports and across time.

The ACCC guidelines state that airport operators must consult them closely on the methodology and contents of the survey, and that a full description of survey methodology is provided to the ACCC with the results. However, consultation would need to go far beyond the current guideline; the requirements as implemented by the ACCC do not appear to be sufficient for ensuring the passenger survey is of sufficient quality. Without clear and consistent guidelines for all airports, and without transparent reporting on methodological detail, the results have little value.

4 Airline survey

The survey of airlines is undertaken by the ACCC. There is very little reference to the airline survey in the ACCC guideline document; only a list of the criteria covered in the survey is provided in an attachment.

The description of the airline survey provided in the Airport Monitoring Report is:

> "The ACCC conducted a survey of airlines to gain information on their perception of the quality of facilities they used at the monitored airports. The facilities and services covered include:
>
> - terminal facilities—including aerobridges, check-in and baggage processing
> - airside facilities—including runways, taxiways, aprons, aircraft gates and ground equipment sites.
>
> Airlines were asked to rate two aspects of these facilities:
>
> - availability— that is, the availability of infrastructure and equipment and the occurrence of delays in gaining access to those facilities
> - standard— that is, the ability of equipment to perform the function intended, the reliability of the equipment and the probability of it breaking down.
>
> The airlines were also asked to rate the airport operator's responsiveness or approach to addressing problems and concerns with the above facilities."
>
> (ACCC Airport Monitoring Report 2008–09, p.14)

Not all domestic airport facilities are relevant to the airline survey:

> "Under the ACCC monitoring regime, airlines are not required to provide survey information for the domestic facilities they operate under Domestic Terminal Leases (DTLs). Melbourne, Perth and Sydney airports lease operating space to Qantas for domestic terminal services. For example, Qantas operates Terminal 1 at Melbourne Airport, Terminal 2 at Perth Airport and Terminal 3 at Sydney Airport. In addition, the majority of Brisbane Airport's domestic terminal is occupied by Qantas and Virgin Blue under DTLs."
>
> (ACCC Airport Monitoring Report 2008–09, p.14)

Other references to the airline survey in the report are those describing the results.

Similar to the passenger survey, the ACCC's website provides an Excel worksheet that is a template for recording survey responses.

With this background information about the airline survey, we consider the key areas for survey evaluation described in Section 2. Many of the important issues raised in the review of the passenger survey in Section 3 are also relevant to the airline survey. To avoid unnecessary repetition, these will be mentioned only briefly here; a full detailed discussion can be found in Section 3.

4.1 The population of interest and sampling frame

The ACCC annual report implies that the population is international and domestic airlines who schedule flights in and out of the five Australian airports of interest. This is general and could be more precise. For example, the report refers to "provision of aeronautical services", but exactly what kinds of services are included should be covered. It is likely that the population of interest is finite in size and relatively stable.

As mentioned in Section 3.2, without an explicit and careful description of the population of interest, an appropriate sampling frame cannot be defined. The sampling frame of the airline survey is not discussed in the Airport Monitoring Report. It is likely to be straightforward to obtain an accurate and complete sample frame of all airlines both international and domestic who schedule flights in and out of the five Australian airports of interest.

The unit of interest in the survey appears to be "airlines"; the survey is concerned with the *perceptions* of airlines. This, in itself, seems highly problematic as it is not clear how an airline can *perceive* the quality of airport services. Airlines are extremely large and complex organisations comprising many and varied staff; the staff can have perceptions of the quality of airport services. It would be reasonable to assess perceptions of airport quality by surveying airline staff rather than "airlines"; in this case, the population units of interest are current staff. However it appears that this is not the way the airline survey is defined.

4.2 Sample selection and sample size

The Airport Monitoring Report does not provide any details regarding the sampling procedures or sample sizes used in the implementation of the airline survey. Given the finite population of airlines flying in and out of Australian airports, it is possible that the airline survey could involve a

census. Indeed, the data reported in the annual report may represent all airlines; this is not specified. A census is not essential; reliable data could be achieved from a sample of airlines.

A simple random sample of airlines from a well-defined sample frame would be possible; however it may be more appropriate to adopt a stratified sampling strategy, stratifying by airline size or frequency of use of Australian airports. This would ensure that the airlines that are the most regular users of Australian airports are more heavily represented in the airline sample. It may be possible to implement a census of the large and regular use airlines and a simple random sample of the small and irregular use airlines.

If the total number of airlines is not large, a census would be appropriate. The size of any sample of airlines will be limited by the finite size of the population. For finite populations, the proportion of the population sampled is important in determining the precision of the survey results.

If the airline survey sampled airline staff (rather than airlines), there are many complex issues that would need to be considered and specified in order to obtain a suitable sample.

4.3 Survey administration and response rates

If, as it appears, a single response is required from each airline then details of the administration procedures are vital. The format and structure of the Excel airline survey template indicates that the ACCC requires one survey to be completed per airline; however, this is not stated explicitly in the monitoring report.

There is no information about who within the airline is responsible for completing the survey. It is not clear who within each airline is provided with the electronic survey template by the ACCC and if and how it gets distributed to the staff member in the best position to complete it. It is not known whether the ACCC provides additional documentation containing specific and detailed instructions to the airlines outlining its expectations regarding these matters. The obligation of airlines to respond to the survey is not discussed.

We do not know if data obtained from an airline represent the responses of one individual within the airline, the consensus decided by a number of staff or the collective responses of many airline staff. Different airline staff will be better qualified to assess some aspects of airport quality of service and less so other aspects. This has implications for quality of survey responses obtained.

It is important that the airlines are provided with guidelines about who is responsible for responding to the survey to ensure consistency across airlines.

There is no description of the methods used by the ACCC in administering the airline survey in the Airport Monitoring Report. The Excel airline survey template implies that the survey is self-administered; it seems likely that the template is provided to the airlines and the survey is completed and returned in electronic format to the ACCC. If the survey is indeed self-administered, the questionnaire must be well-designed and easy to follow, and the wording of question must be very clear. This will be discussed further below.

The quality of responses as well as the completeness of responses to the airline survey is likely to depend on who receives the survey. There is no mention of response rates or what is considered acceptable by the ACCC in the Airport Monitoring Report. Given the finite population and the method of survey administration, airlines not responding should be able to be identified. The ACCC should be able to follow-up non-responders to ensure a high response rate. Information about the airlines such as size, frequency of use of Australian airports and whether they provide domestic or international services should be readily available, and therefore it should be possible to investigate the potential for non-response bias. As the report currently stands, we have no information about the relevance of these issues to the airline survey.

4.4 Questionnaire design and quality of responses

The ACCC guidelines and monitoring report specify the scope of issues to be covered in the airline survey. The airline survey criteria for monitoring are detailed in Attachment A of the guideline document, and the measures are also listed in Appendix A2.1.3 of the Airport Monitoring Report.

The Excel airline survey template on the ACCC website appears to be the survey tool distributed to sampled airlines for completion and submission to the ACCC. The file contains six worksheets: the first labelled "Airline details" requires the airline name, date and airline survey co-ordinator contact details to be entered. In addition, the ACCC provides a set of notes and definitions to assist airlines with completing the templates. The remaining five worksheets are identical in format and require survey responses to be entered by the airlines for Adelaide airport, Brisbane airport, Melbourne airport, Perth airport and Sydney airport respectively.

The airline survey template appears to be reasonably well-designed. Useful airline information is collected for each airport such as any subsidiary airlines, the type of service (domestic, international or freight/cargo) and the average numbers of arriving and departing flights and passengers per week.

Survey respondents are asked to assess aspects of airport quality of service and enter a * in cells to indicate ratings on an ordinal scale of 1 to 5 (where 1 = Very poor, 2 = Poor, 3 = Satisfactory, 4 = Good and 5 = Excellent). 'Not applicable' responses are also allowed and are encouraged when questions relate to facilities/services that are non-existent at an airport or are not provided by the airport operator. Respondents are also encouraged to explain 'not applicable' responses in more detail in the cells requesting additional qualitative comments.

The structure and layout of the survey items appear to be clear and all important or potentially ambiguous terms are explicitly defined in the airline details worksheet. The survey also appears to be reasonable in length.

There are some features of the questionnaire design that have important implications for the quality of the survey responses. The airline survey is not anonymous; airlines are required to provide their name on the survey form as well as the names of the staff in points of contact at each airport. This means that airlines are aware that they can be directly linked to their responses which may influence the accuracy and quality of the ratings they provide.

In the ACCC guideline document it states:

> "In addition, airline surveys and … 'whole-of-government' surveys are to be reviewed and submitted by the relevant airline's and government agency's head office respectively. The ACCC considers that this will allow for the results to account for commercial negotiations and reduce the potential for bias. Where an airline or government agency gives a rating of below satisfactory, they must support this with commentary detailing the complaint and steps they have taken to inform the airport operator of their complaint."
>
> (ACCC guidelines, p.8-9)

This additional ACCC requirement aims to discourage airlines from negatively rating airports unfairly in order to gain advantage in future commercial negotiations. However, it is possible that the additional requirement for below satisfactory ratings might discourage respondents from making such a rating in situations where it might reasonably apply.

4.5 Reporting of the results

The measurement scales used for the airline survey are ordered categories assigned arbitrary values 1 to 5. As discussed in detail in Section 3.8, although the assignment of numerical values to categories is quite a common practice, it is inappropriate, without very strong assumptions about the nature of the ordinal scale.

The results of the airline survey are reported in the Airport Monitoring Report in a very similar way to the passenger survey results. Estimated values of population characteristics are reported without any indication of variability or error and no formal statistical inferences are reported when making comparative statements between airports and across survey years. If a census is achieved each year, there may be no need for reporting indicators of variability. If measures of variability are needed, the statistical methods used should account for the relatively small finite population sizes; this situation is different from the sampling of passengers where the population is effectively "infinite", meaning it is extremely large.

In combining survey responses across the responding airlines, the ACCC states that:

> "Ratings given by airlines were averaged (with equal weights) to give an overall rating for each facility at each airport."
>
> (ACCC Airport Monitoring Report 2008–09, p.14)

Applying equal weights to airlines in calculating the airline quality of service rating may not be the most appropriate approach. Airlines are likely to vary considerably in their frequency of use of certain airports. For example, some airlines may fly in and out multiple times per day while others may schedule only a few flights per week. Weighting responses based on the frequency of use could be considered.

As mentioned above, it is inappropriate to assign numerical values to categories without very strong assumptions about the nature of the ordinal scale. There are a number of alternate and valid ways of interpreting the measurement scales. The Airport Monitoring Report states that:

> "… airline quality ratings are used as indicators of whether an airport has provided quality above a minimum efficient level—that is, an average rating of at least satisfactory (3.0) by user airlines. In a competitive market, it is expected that a firm would be unable to sustain a level of quality that is below satisfactory as it would lose its customers to rivals."
>
> (ACCC Airport Monitoring Report 2008–09, p.33)

If a rating of at least satisfactory is interpreted as a minimum efficient level for airport quality of service performance, then a valid way of summarising the measurement scale would be to calculate the proportion (or percentage) of responding airlines who rated each airport as satisfactory or better on each service criteria. An overall measure of the quality of service of an airport as rated by an airline could be obtained by summing up the number of criteria for which the airline rated the airport as satisfactory or better. This is an example of a clearly defined and transparent summary measure.

An overall airline rating of the quality of service is reported but as elsewhere there is no explanation about how this overall rating is obtained from the individual survey questions. The airline survey results are also presented graphically using line graphs and bar charts that plot the average rating using labels of Very poor, Poor, Satisfactory, Good and Excellent equally spaced on the vertical axis. A detailed discussion of the problems associated with each of these methods of reporting in relation to the passenger survey can be found in Section 3.8.

4.6 Overall quality of the survey

As the airline survey is conducted by a single agency, the issues of consistency in design and implementation that arise for the passenger survey are not relevant. However, there is not enough information in the guidelines and the Airport Monitoring Report to assess the quality and appropriateness of the statistical methodology used by the ACCC in implementing the airline survey. Without such information, the overall quality of the survey results cannot be judged.

However, as we suggested for the passenger survey, the airline survey results do not have a clear interpretation. The ACCC may have implemented a valid and appropriate methodology, but without transparent reporting on methodological detail, the results have little value. Most importantly, readers of the report need to know who within the airlines were targeted by this survey and how it is ensured that the survey responses accurately reflected the "perceptions" of the airline, as well as the response rate and sample size. A naïve reader might assume, not unreasonably, that airline staff had been sampled. A naïve reader might infer that there were strong differences between airports or over time without considering the statistical uncertainty in the summary measures presented.

5 Border agency survey

The survey of border agencies is undertaken by the ACCC. There are three border agencies of interest: the Australian Customs and Border Protection Service (AC&BPS), the Department of Immigration and Citizenship (DIAC) and the Australian Quarantine and Inspection Service (AQIS). There is very little reference to the border agency survey in the ACCC guideline document; only a list of the criteria covered in the survey is provided in an attachment.

The description of the border protection survey provided in the Airport Monitoring Report is:

> "For the 2007–08 monitoring report onwards, the AC&BPS has agreed to coordinate a government survey that encompasses the views of the AC&BPS, the Department of Immigration and Citizenship and the Australian Quarantine and Inspection Service to provide a wider perspective of government inspection services."
>
> (ACCC Airport Monitoring Report 2008–09, p.15)

> "The survey of the three agencies sought ratings of facilities in the following areas provided by airports:
>
> - arrivals (immigration)—adequacy of areas for circulation and queuing, signage, lighting, desks and passenger facilities (for example, seating and toilets)
> - arrivals (baggage inspection or examination area)—adequacy of space to avoid congestion, signage, provision for passenger privacy, appropriate access and security, and passenger and inspection facilities
> - departures (immigration)—adequacy of circulation space to avoid congestion, signage and appropriate provision of desks.
>
> The agencies were asked to give separate ratings for each area for:
>
> - adequacy—the space and facilities made available for its operations, covering the amount of space provided and the likelihood of congestion and delays in passenger processing
> - standard and condition in which areas are generally maintained. The border agencies were also asked to rate the airport operator's responsiveness or approach to addressing problems and concerns with the above facilities."
>
> (ACCC Airport Monitoring Report 2008–09, p.16)

Other references to the border agency survey in the report are those describing the results.

Unlike the passenger survey and the airline survey, there is no Excel template available on the ACCC's website for recording responses to the border agency survey.

We now consider the key areas for survey evaluation of the border agency survey, focussing on particular issues that need elaboration for this survey.

5.1 Population and sampling issues

The border agency survey aims to encompass the views of AC&BPS, DIAC and AQIS; this implies that the population of interest is these three border agency organisations. However as is the case with airlines, border agencies are large organisations comprising many and varied staff. It is not clear from the Airport Monitoring Report who within the agencies is targeted by the survey and how it is ensured that the survey responses reflect the "perceptions" of the entire organisation. It is not clear whether one or more than one set of survey responses is obtained from each agency. Again, we need to know if an agency provides a single response or if the survey targets staff within an agency.

If this survey does only involve three responses from three agencies, it is somewhat confusing to call it a "survey". Sampling procedures and sample sizes are not relevant as all agencies should be included.

If the survey targets agency staff, important details about the sample frame, sampling procedures and sample sizes need to be made explicit. Such a survey would be complex to design and manage.

5.2 Survey administration and questionnaire

The Airport Monitoring Report states:

> From 2007–08, the border agency survey encompassed the views of the AC&BPS, the Department of Immigration and Citizenship (DIAC) and the Australian Quarantine Inspection Service (AQIS). The results were collated by the AC&BPS.
>
> (ACCC Airport Monitoring Report 2008–09, p.16)

If only one response is obtained from each agency, it is important to know how the response was obtained and who contributed to it. Presumably "collation" in the statement above includes data collection; it should be clarified if this is a formal or an informal process. It is not clear who within each border agency is provided with the survey and if and how it gets distributed to the staff in the best position to complete it. As with the airline

survey, it is not known whether the ACCC provides additional documentation containing specific and detailed instructions to the agencies outlining its expectations regarding these matters.

Given the finite population of size three, it is likely that all agencies would be expected to participate and that all agencies have responded; this should be stated clearly in the report. If any of the agencies did not participate, this should be explicitly stated. Non-response from any one border agency would result in potential for response bias.

In all surveys discussed here, the scope of issues to be covered is clearly articulated in ACCC documentation. The border agency survey criteria for monitoring are detailed in Attachment A of the guideline document, however the measures are not listed in the appendix of the Airport Monitoring Report where lists of the measures for the passenger and airline surveys are provided.

There is no information about the exact wording of questions in the border agency survey and there is also no information regarding survey format or design. The measurement scales are not defined however based on the reporting of results it appears that the five-point ordered scale with arbitrarily assigned numerical values used in the passenger and airline surveys is also used in the border agency survey.

The issues of anonymity, who within an organisation is responsible for completing the survey and potentially unfair negative ratings discussed in detail in Section 4.5 in regard to the airline survey, are also relevant for the border agency survey.

5.3 Reporting of the results

The results of the border agency survey are reported in the Airport Monitoring Report in a very similar way to the passenger survey and airline survey results. Numerical values are arbitrarily assigned to an ordered categorical scale and results are presented graphically using line graphs and bar charts that plot the average rating using labels of Very poor, Poor, Satisfactory, Good and Excellent equally spaced on the vertical axis.

Estimated values of population characteristics are reported without any indication of variability or error and no formal statistical inferences are reported when making comparative statements between airports and across survey years. As for the airline survey, it is not clear if such measures are

required or not; this depends on whether all agencies have been included or not.

As elsewhere, an overall rating of the quality of service is reported without an explanation of its derivation or meaning. A detailed discussion of the problems associated with each of these methods of reporting in relation to the passenger survey can be found in Section 3.8.

5.4 Overall quality of the survey

The design and method for "surveying" border agencies may be very good; however with the limited information publically available about the survey no judgment about the survey quality can be made. Readers of the ACCC's report need to know if the border agency survey is actually a census of three agencies where a single response is obtained from each agency or if it represents a survey of individuals working within the relevant agencies. Precise descriptions and definitions should appear in the report to avoid such ambiguities. Hence, the border agency survey results do not have a clear interpretation.

6 Other issues

The aggregation of results from the three surveys and the objective measures was briefly discussed in Section 3.7:

> "The ACCC monitors the airports' dealings with airlines as well as other stakeholders, such as passengers and border agencies. The ACCC determines an overall rating of quality of service, taking into consideration a range of quality of service indicators, including survey responses from all of these airport users."
>
> (ACCC Airport Monitoring Report 2008–09, p.ix)

It is unclear how this aggregated result is obtained. For example, are the responses from the different stakeholders equally weighted? The interpretation of this kind of measure rests on the way it is constructed. The measure combines information from different types of populations, and as such, has an ambiguous frame of reference. The error on the measures taken from different sources is likely to be very different; it is not obvious how the measures of variability from the different sources would be combined for the overall measure.

The ACC report notes that AirServices Australia data was collected for four of the five airports – Adelaide was not included. It should be clearly explained how this was dealt with in creating the aggregated quality of service scores.

The ACCC ranks airports on this basis:

> "Taking all users' perceptions into consideration as well as a range of objective indicators about quality of service, the ACCC ranks the airports relative to each other."
>
> (ACCC Airport Monitoring Report 2008–09, p.ix)

The ranking may be useful (leaving aside the issue of the meaning of the overall score for the moment) but if there is a large amount of error in the overall ratings, the rankings may be meaningless. Airports with adjacent ranks could have minimal differences on the overall rating or relatively large differences.

The ACCC report compares measures obtained from different sources:

> "Interestingly, passengers' perceptions were, in general, higher than the overall ratings of quality of service for the monitored airports. While airports provide much of the services and facilities at the airports, the airlines and border protection agencies also provide services that may influence the passenger experience. Therefore, airlines' and border agencies' ratings of the airports can vary significantly from passengers' perceptions and

> provide another indicator of the airports' quality of service." (ACCC Airport Monitoring Report 2008–09, p.61)

Any comparative claim like that made above should take into account the precision of the estimates involved. If the precision of any of the measures mentioned above is poor, claims that a difference is important or significant cannot be substantiated. Again, the precision of data from these very different sources is likely to be very different. At present the ACCC report does not provide any basis for making comparative claims.