

Quality assurance of administrative data used in air transport statistics

Investigation of the administrative data sources used in the production of short-term economic output indicators by ONS's National Accounts and Economic Statistics Group.

Contact:
Karen Carter and Glyn Rice-
Mundy
stoi.development@ons.gsi.gov.
uk
+44 (0)1633 455272

Release date:
16 April 2018

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [Quality assurance of administrative data \(QAAD\) assessment](#)
3. [Areas of quality assurance of administrative data \(QAAD\)](#)
4. [Summary](#)

1 . Introduction

1.1 Background

The National Accounts and Economic Statistics (NAES) within the Office for National Statistics (ONS) collects data on the air transport industry from the Civil Aviation Authority (CAA) UK [airline data](#) and [airport data](#) releases. These data form one source in the calculation of short-term economic output indicators, namely gross domestic product (GDP(O)) and the Index of Services (IoS) for the UK.

This report outlines the process data take from initial collection through to the output of the release. It identifies potential risks in data quality and accuracy as well as details of how those risks are mitigated.

This report forms the latest in a series of quality assurance of administrative data (QAAD) reports produced by NAES to investigate the administrative data sources we use in the production of short-term economic output indicators as set out by the [UK Statistics Authority](#). As such this report focuses on our administrative data use for the air transport industry (SIC 51) only. Separate industries where we utilise administrative data will be considered in other QAAD reports in the series.

Further information relating to quality and methodology for the short-term economic output indicators can be found in our [Gross domestic product, preliminary estimate](#) and [Index of Services QMI](#).

1.2 Standard Industrial Classification (SIC) overview

The air transport industry covers all activities under UK Standard Industrial Classification: SIC 2007 division 51. This division includes the transport of passengers or freight by air or space and is part of Section H Transportation and storage.

Based on the [UK SIC 2007](#) the industry is classified to two groups:

51.1 – Passenger air transport

51.2 – Freight air transport and space transport

According to the Inter-Departmental Business Register (IDBR)¹, there were 785 enterprises classified under division 51 in March 2016. This is a decrease of approximately 25 enterprises (negative 3.0%) from the previous year (March 2015).

Most enterprises within division 51 were allocated to 51.1 Passenger air transport, which equates to 585 enterprises (74.5% of the total division).

Notes for: Introduction

1. The [Inter-Departmental Business Register \(IDBR\)](#) is a comprehensive list of UK businesses that is used by government for statistical purposes. It provides the main sampling frame for business surveys carried out by ONS and other government departments. It is also an important data source for analyses of business activity.

2 . Quality assurance of administrative data (QAAD) assessment

2.1 UK Statistics Authority QAAD toolkit

The assessment of our administrative data sources has been carried out in accordance with the [UK Statistics Authority Administrative Data Quality Assurance Toolkit](#).

Each administrative data source investigated has been evaluated according to the toolkit's risk and profile matrix (Table 1), reflecting the level of risk to data quality and the public interest profile of the statistics.

Table 1: UK Statistics Authority quality assurance of administrative data (QAAD) risk and profile matrix

Table 1: UK Statistics Authority quality assurance of administrative data (QAAD) risk and profile matrix

Level of risk of Quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of a low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
High	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

Source: Office for National Statistics

The toolkit outlines four specific areas for assurance and the rest of this report will focus on these areas in turn. These are:

- operational context and administrative data collection
- communication with data supply partners
- quality assurance principles, standards and checks applied by data suppliers
- producer's quality assurance investigations and documentation

In the assurance of our data source, we have chosen to give a separate risk and profile matrix score (Table 1) for each of the four areas of assurance. This will allow us to focus our investigatory efforts on areas of particular risk or interest to our users (Table 2).

2.2 Assessment and justification against the QAAD risk and profile matrix

Table 2: QAAD risk and profile matrix assessment of administrative data used to measure activities of air transport

	Low	Medium	High
	A1	A2	A3
Operational context and administrative data collection		A2	
Communication with data supply partners		A2	
Quality assurance principles, standards and checks by data supplier		A2	
Producers quality assurance investigations and documentation		A2	

Source: Office for National Statistics

The risk of quality concern and public interest has been set as “medium” due to the contribution that the air transport statistics feed into the Index of Services (0.5%) and gross domestic product (0.4%). As such, a score of A2 (medium assurance) is deemed appropriate for this data source.

All scoring was carried out by National Accounts and Economic Statistics (NAES) based on the level of risk of the data and interest of our users. Results for each area of assurance for air transport are shown in Table 2. If you feel that this report does not adequately provide this level of assurance or you have any other feedback, please contact stoi.development@ons.gov.uk with your concerns.

3 . Areas of quality assurance of administrative data (QAAD)

3.1 Operational context and administrative data collection (QAAD matrix score A2)

This relates to the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being compiled and the factors that might increase the risks to the quality of the administrative data.

The [Civil Aviation Authority](#) (CAA) is the UK’s independent specialist aviation regulator and is a public corporation of the Department for Transport. CAA is responsible for collecting and analysing a range of statistics related to UK aviation; much of these data are freely available to the public. CAA funding comes entirely from revenue from charges to the air transport industry, which it regulates, and those who use their services such as air operators requesting permits to fly.

As the CAA is independent of government, the statistics they produce are not subject to the Statistics and Registration Services Act 2007. However, the [Code of Practice for Statistics](#) has been used to provide guidance on best practice for the use of these data.

National Accounts and Economic Statistics (NAES) collect both airport and airline data from CAA publications; these data cover both passenger air transport (SIC 51.1) and freight air transport (SIC 51.2) usage. NAES uses these data as a volume measure for the air transport industry, which forms part of the Index of Services (IoS) that feeds in to gross domestic product (GDP(O)).

3.1.1 Airline data

NAES download five of the monthly [airline data](#) tables published by CAA on their website, for use as volume measures for passenger miles and miles per tonne of freight for scheduled and non-scheduled air services. The tables used are as follows:

- 4.1 All scheduled services
- 4.2 European Economic Area (EEA) international scheduled services
- 4.3 Other international scheduled services
- 4.4 Domestic scheduled services
- 5.1 All non-scheduled services

CAA's reporting criteria states that holders of a UK Air Operator Certificate (AOC) are required by law to report monthly data on each route served. Data is required under [Section 84 \(1\) of the Civil Aviation Act 1982](#) to carry out their duties and meet the requirements of the International Civil Aviation Organization and Eurostat. Using this legislation, the CAA collects monthly data from airlines with aircraft having a Maximum Take Off Mass (MTOM) above 40 tonnes, and/or an A Type Operating licence, or an Air Transport Licence. These are included in the tables published by CAA and used by NAES.

Some other chargeable air movements, for example, charter services performed by aircraft below 15 tonnes (MTOM) and services such as training flights and staff ferry flights, are not included in published tables. However, these are likely to be a small percentage of air movement.

CAA publishes airline data tables both monthly and annually; finalised data are available usually within 11 weeks after the data month.

3.1.2 Airport data

Similarly to airline data (in section 3.1.1), NAES downloads airport data tables from the CAA website for data on passenger numbers. The data table used, Table 9 – Terminal and transit passengers, includes total passengers on board the aircraft, apart from working crew.

The airport authorities supply the data to CAA for this release, although the data originate from airlines and handling agents such as SERVISAIR, who report to the airport authorities. These data are collected from mainland airports that handle scheduled, non-scheduled and major charter services. Airports handling certain levels of traffic are also included in the coverage of these reports. In total, this comes to approximately 60 airports. These data are published both monthly and annually. Provisional data are released two weeks after the end of the data month and these are then updated weekly until all the data have been received and validated. Provisional data are clearly indicated by a prominent note on the relevant data page. The data are usually finalised within seven weeks, resulting in a nine-week gap from the end of the data month and finalised publication. NAES uses a series of forecasting tools to overcome these arrears; this methodology is based on previous year trends and user expertise (more details are included in section 3.4).

3.1.3 CAA revisions policy

CAA balances timeliness, the requirements of users and accuracy by publishing some data provisionally then updating the dataset until the data are complete, as mentioned in section 3.1.2. These datasets are clearly marked as provisional and a prominent note indicating what data are missing is displayed on the data page. CAA provides users with clear information concerning large revisions to the dataset. Users are notified with varying urgency depending on the severity of the revision or error:

- minor revisions – corrections feature in the following publication
- major revisions – electronic publication amended as soon as possible, alerts are placed on the webpage, and subscribed users are notified of the publication of revised data.

3.1.4 Potential challenges

CAA airline data is not a census of air transport; it contains only the airlines that meet the CAA reporting criteria. For example, in the CAA notes, it identifies that Isles of Scilly airbus no longer meets the reporting criteria – this is because it has not completed full returns to the CAA and so it is not included in the published data tables. However, this is still an airline that creates output in the air transport industry, and therefore should be included in IoS. Similarly, these publications do not include certain smaller aircraft performing charter services, which again are services that should be included; the CAA are not aware of the percentage of under coverage from these exceptions but it is believed to be small. However, CAA do report all the larger air movements from all the major airports, capturing all long-haul flights. This is very significant to the IoS results and there is not currently a more comprehensive coverage of the industry.

The strengths of this method are:

- impartial statistics, created using the [Code of Practice for Statistics](#) to provide guidance on best practice
- comprehensive coverage of the air transport industry
- data is required under the Civil Aviation Act to meet the requirements of International Civil Aviation Organization and Eurostat
- data collected monthly provisional data and revisions

The weaknesses are:

- not a census of all airlines, however this is likely to be a small percentage of air movement
- not a census of all airports excluding some smaller airports
- time lag between the data period and publication means that estimation must be used to overcome these arrears
- not currently a subscribed user to be updated about major revisions

Next steps for NAES are to arrange with the CAA to become “subscribed users”, ensuring notification of any major revisions. It will also confirm the potential scale of impact of missing data and then investigate estimate for services not covered in the data published by CAA in the airline and airport data tables.

3.2 Communication with data supply partners (QAAD matrix score A2)

This relates to the need to maintain effective relationships with suppliers (through written agreements such as service level agreements or memoranda of understanding), which includes change management processes and the consideration of statistical needs when changes are being made to relevant administrative systems.

3.2.1 CAA communication with data providers

The Civil Aviation Authority (CAA) has expressed that it has a good working relationship with its airport and airline data suppliers and maintains regular contact and a high degree of engagement in a number of ways including:

- a published set of detailed user guides and guidance notes helping suppliers to understand and access their system
- webinar links, where suppliers can view 30-minute recordings on YouTube on specific subjects such as: how to submit data, what happens if your data is rejected and questions and answers from the webinar
- monthly contact through email and telephone to query data submission and data changes
- pre-booked face-to-face meetings to overcome barriers such as suppliers experiencing major difficulties providing data (however, CAA say this service is not used very often as most issues are solved via the telephone)
- additional help and guidance in place through a member of the CAA business intelligence team to answer complex queries
- up-to-date frequently asked questions page, including airport and airline data changes

CAA says that contact with data providers is also useful as these tend to be a subset of their final users and feedback from them on their published airport and airline reports is welcomed. They review every comment they receive and, if significant issues are found with their publications, reports will be amended and re-published.

3.2.2 NAES communication with CAA

National Accounts and Economic Statistics (NAES) do not have a formal working agreement with the Civil Aviation Authority (CAA); this is because the data that NAES use are free of charge and publicly accessible through the CAA website. The logistics of implementing and actively managing formal arrangements are considered both prohibitive and unnecessary for this data, considering the weight of the Index of Services (IoS) and gross domestic product (GDP (O)) it comprises.

Previously NAES has had difficulties in contacting CAA, with CAA stating that they are unable to respond to individual queries. More recently, however, NAES has established more frequent contact with CAA, for example, concerning a query raised during the NAES quality assurance process, where some data appeared too high. CAA then acknowledged and corrected this mistake. Further contact has been established through the writing of this report and we intend to maintain this level of communication.

As discussed in section 3.1.3, CAA clearly communicates major revisions on its website to subscribed users. A potential issue identified with this previous lack of relationship with CAA is that NAES would not be contacted directly or notified in the case of major issues with the data, or changes to the methodology. Therefore, NAES would benefit from becoming a subscribed user of data so that they can be alerted in the event of changes to the data, and risks to the data would be reduced. Currently the business as usual team pay close attention to the CAA website to ensure they are fully up to date with revisions and methodology changes. Although CAA publishes these, NAES would benefit from being informed in advance to make any adaptations needed to compensate for changes.

The strengths of this method are:

- CAA regularly communicates with data providers
- CAA uses a range of communication methods with data providers
- additional help for data suppliers to answer complex queries
- CAA publish methodology changes on their website
- CAA produce detailed user guides and guidance notes for suppliers
- data consistently available and clearly communicated through CAA website

However, the weaknesses are considered to be:

- no formal contact between NAES and CAA
- past difficulties in contacting CAA to query data
- NAES not subscribed user of CAA data and so changes to methodologies and policies are not communicated directly to NAES

The next steps are to maintain the newly established communication links developed during the writing of this report. NAES would benefit from being informed in advance of methodology and policy changes.

3.3 Quality assurance principles, standards and checks by data supplier (QAAD matrix score A2)

This relates to the validation checks and procedures undertaken by the data supplier, any process of audit of the operational system and any steps taken to determine the accuracy of the administrative data.

From April 2016, CAA introduced a new system for reporting and processing airport and airline data. Two main development changes were introduced to improve the quality of outputs, which were:

- standardised templates for submitting data
- greater responsibility for data suppliers and automated prompts for correcting any incomplete or incorrect data submissions through the secure portal, increasing timely record completeness and accuracy

New templates are completed by the airport and airlines data providers. They are uploaded into the secure online portal, which performs high-level automatic checks against many mandatory fields to ensure input data is correct. This includes:

- incorrect codes
- data format
- UK registration matches the reported aircraft type and owner
- validating that a specific aircraft can travel the distance reported

Data suppliers then receive an email notification of submission status if rejected and reasons are provided asking them to correct the information and resubmit.

As a further check, CAA also check data against aircraft, operators and airports reference data, which is held on the portal internally within the CAA. Once the data has been accepted through the automated checks, the experience business intelligence team performs final validation checks on the data. Some of the checks are prompted as messages through the online portal and others by team members undertaking various integrity checks based on their knowledge of operating patterns and the industry. These include differences in percentage terms based on previous months, same month and the previous year. Reports are also checked for accuracy to ensure no other unexpected changes have occurred.

The strengths of this method are:

- new standardised templates to improve submissions
- secure online portal for upload including comprehensive automated checks
- alert to suppliers to submit data or incorrect data
- further validation checks completed by experience team
- comparisons against reference data

3.4 Producer's quality assurance investigations and documentation (QAAD matrix score A2)

This relates to the quality assurance conducted by the statistical producer, including corroboration against other data sources.

NAES has specific desk instructions for collecting and processing air transport data; the most recent update of these instructions was in 2016. The processes used for compiling these data have not changed since this update. NAES downloads data from the six CAA data tables outlined in section 3.1 and 3.1.2. These data are then pasted into the relevant columns in the Excel spreadsheet used by NAES. The data is then sense checked against similar periods over previous months and years to ensure the data is in line with what would be expected. Comparisons are checked for monthly growth rates, year-on-year trends and revisions.

As mentioned in section 3.2, this quality assurance process led to the discovery of an error in the data published by CAA, who then updated their records accordingly. This error was spotted prior to the publication of IoS; however, CAA did not correct the error until after the publication. Some data for this period were estimated as a result.

The data are stored in a CSV file format and each historic file is saved onto the network drive to form an audit trail. The CSV file for the latest period is then uploaded to an internal ONS system where missing data owing to the time lag between publication by CAA and the data month is forecasted. In addition, seasonal adjustment is applied at this stage.

This system can forecast missing data, to compensate for the time lag between the data month and the publication of CAA data, and create an estimate for the months that are missing. The internal system forecasts the data at the lowest industry classification level, based on specification files inputted by the time-series analysis branch. These files specify which model is the best to use for forecasting for this series. In the case of CAA data, the forecast is carried out using an X13-ARIMA-SEATS model that forecasts missing data based on previous years' returns. In the event of missing data from major airports, for example, Heathrow and major airlines like British Airways, NAES does not use the newest data but forecasts the most recently reliable data from CAA. Finally, the final output is quality assured and signed off by two senior managers within the team.

NAES has no specific customer for air transport statistics, as this makes up just one component of the IoS in which it is published. NAES has not received any feedback to date on the statistics they produce relating directly to air transport. There have also been no significant revisions to this industry in recent years. This industry is of medium public interest, particularly owing to its IoS weighting (0.5%) and gross domestic product weighting of 0.4%.

The strengths of this method are:

- established quality assurance in place lead to the discovery of an error in CAA data
- data are compared by looking at growth rates year-on-year trends
- final output is quality assured by two senior managers

The weaknesses are considered to be:

- do not estimate for airlines not covered by CAA statistics – these are likely to be a small percentage of air movement
- use of copying and pasting in Excel
- identified error not corrected by CAA in time for release and estimate

Next steps for NAES are to improve automated procedures to populate internal systems with updates.

4 . Summary

[Process map showing the overall flow of Airline and Airport from suppliers to ONS](#)

In investigating the administrative source for the activities of air transport, NAES considers the main strengths of the data for our purpose to be:

- largest coverage currently available
- impartial statistics based on the Code of Practice for Statistics
- detailed knowledge of the subject by Civil Aviation Authority
- comprehensive quality assurance principles in place

We believe the current limitations of this data source are:

- lack of communication between CAA and NAES
- these data are not a census of the air industry and do not cover all chargeable air movements
- time lag between publication of CAA data and the needs of NAES

We will be taking next steps to investigate and address these limitations in our data sources. These steps will be communicated to users in the future QAAD report updates for this topic.

NAES consider this data source to be sufficiently robust to fulfil the requirements of an A2 assurance rating. This is despite the limitations outlined and taking into account the small contribution that the air transport statistics make to the Index of Services (0.5%) and gross domestic product (GDP(O)) (0.4%).