Quality Assurance Process for NRPS

Background

The UK Statistics /Authority (UKSA) approach to quality assurance of administrative data is given below:

“Quality assurance of administrative data is more than simply checking that the figures add up. It is an ongoing, iterative process to assess the data’s fitness to serve their purpose. It covers the entire statistical production process and involves monitoring data quality over time and reporting on variations in that quality.

Post collection quality assurance methods, such as data validation, are an important part of the quality assurance process, but can be of limited value if the underlying data are of poor quality. The Authority encourages the application of critical judgment of the underlying data from administrative systems before the data are extracted for supply into the statistical production process. As with survey data, producers need to: investigate the administrative data to identify errors, uncertainty and potential bias in the data; make efforts to understand why these errors occur and to manage or, if possible, eliminate them; and communicate to users how these could affect the statistics and their use.”

From the UKSA standard, summaries of each of the three possible assessment levels are detailed below:

> A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements
>
> A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance
>
> A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit

These three levels of assurance are applied across a range of four areas relating to administrative data provided for producing official statistics as outlined below:

1. Operational context & administrative data collection
2. Communication with data supply partners
3. QA principles, standards and checks applied by data suppliers
4. Producer's QA investigations & documentation

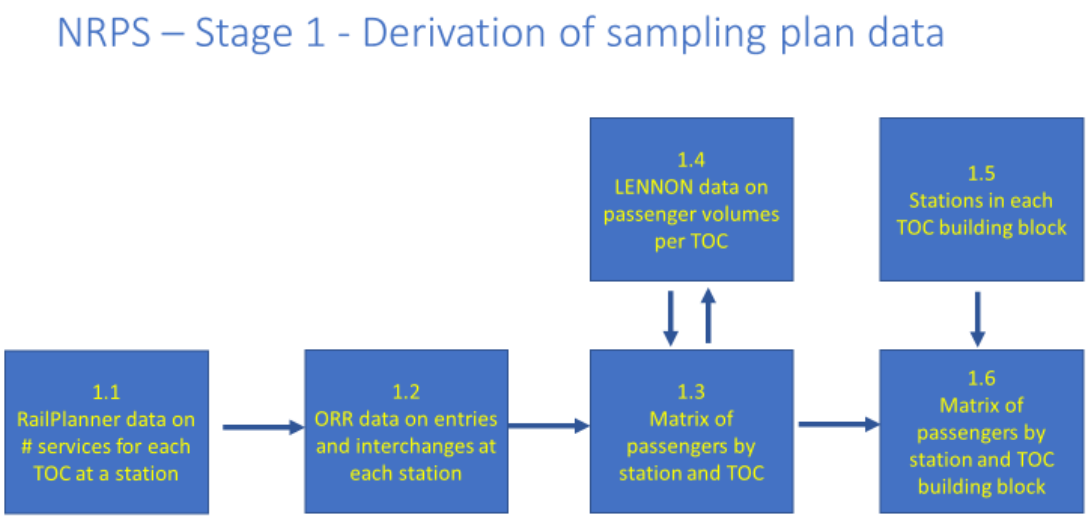
With these guidelines in mind, we have produced this documentation of the NRPS process, from the use of the external data which helps define the sampling plan through to the checks that are made on the data produced. Our aim is to be both transparent (so users of the NRPS statistics can see how information is used in their production) and proportionate (so we don't go over lots of ground that others have already covered). Recommended QA tasks will be built into NRPS processes and checked each wave.

Stage 1 – derivation of sampling plan data

NRPS involves distributing questionnaires to passengers either:

- At stations, immediately prior to their journey
- On train, during their journey

The specific services that use each of these options are defined each wave in the NRPS Technical Guide. Deciding which stations and trains to sample is a result of the process shown below:



The objective of this first stage is to create a matrix of TOC building blocks by station, which lists the estimated annual passenger volume for each station in each TOC building block and can therefore act as a sampling frame for NRPS.

The initial input data is a matrix of station by TOC, showing how many services each TOC operates at each station. This comes from an analysis of RailPlanner data (box 1 above), usually for four weeks in February. A complete number of weeks is chosen to ensure each day of the week is represented at its right level; February is chosen as a month in the year with (normally) four complete weeks, so that there is no issue over which particular days are chosen.

For each station, the number of passengers shown in the ORR data (box 2 above) is split between TOCs calling at that station according to the split of services calling at that station; this implicitly assumes that each service carries the same number of passengers (but as will be seen many other iterations are made to this data which means it is not particularly crucial).

The result of this process is an initial matrix of passenger numbers by station and TOC (box 3 above). The data for each TOC is aggregated and compared to the annual passenger numbers for the TOC as estimated by LENNON (box 4 above). A factor is computed which is the TOC total from LENNON data divided by the TOC

total in box 3 and this factor is applied to each station at which the TOC calls, generating a revised matrix of passenger numbers by station and TOC. This data is aggregated for each station and a factor computed which is the original station total divided by the updated figure; this factor is applied to each station to again update the matrix.

This process continues until the factors converge at a point when the TOC totals agree with the LENNON original and the station totals may vary a little from the original ORR data but are not materially different from the factors produced in the previous station iteration. This is the final matrix of passenger numbers by station and TOC.

TOCs are divided into building blocks, which are usually routes (comprising several stations) and occasionally stations (box 5 above). For a TOC with route based building blocks a particular station may appear in several building blocks; if so, the passenger numbers for that station are divided equally between the building blocks in which is appears. For a station which appears in only one building block, its total passenger number is assigned to that building block. For station based building blocks, all stations are treated this way.

At the end of this process the matrix of passenger volumes for each station and each TOC has been expanded to a matrix of passenger volumes for each station and each TOC building block (box 6 above). Passenger numbers for each TOC building block can be computed by aggregating the data for each station in the building block.

This process raises some QA questions, which should be addressed each wave:

1.1 How does the Rail Planner data compare with the same data from the previous occasion on which this exercise was undertaken? Are there stations where the TOC split has changed a lot? Are there TOCs where the station split has changed a lot? If so, is there supporting evidence to substantiate these changes? Confirm that four weeks in February was used to create the data

1.2 How does the ORR data compare with the same data from the previous occasion? [Steer, the contractor that produces the ORR data already provides extensive documentation on any significant changes in passenger volumes so comment on any significant changes shown in their report]. Confirm the passenger volumes were created by adding entries to interchanges, to form the estimate of passengers boarding at each station

1.3 How does the final matrix of passenger volumes by station and TOC compare with the same data from the previous occasion? Highlight any significant changes and provide comment on why these may have occurred?

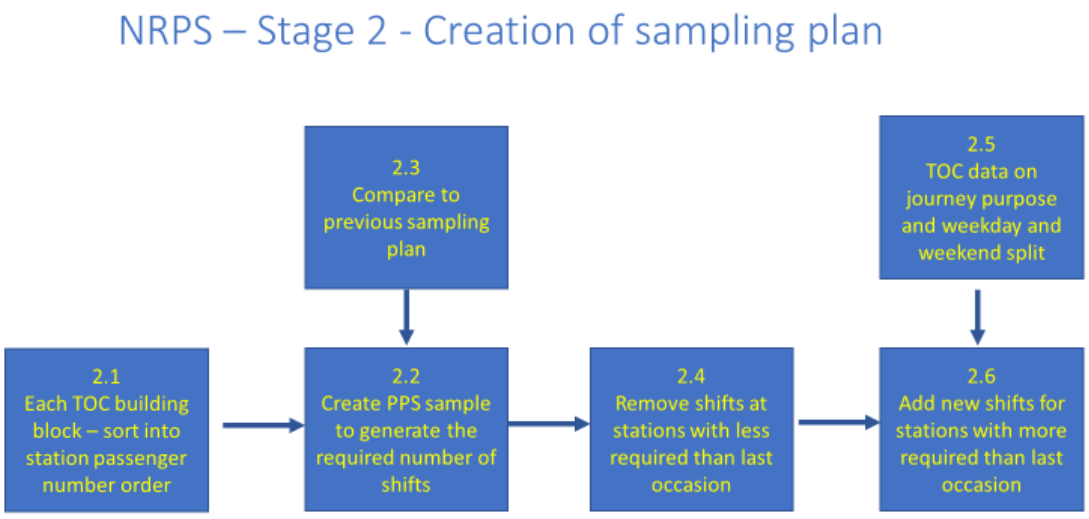
1.4 How does the LENNON data compare with the same data from the previous occasion? Is there any evidence to support significant increases or decreases?

1.5 How does the list of stations in each building block compare to the same data from the previous occasion? Provide proof of changes signed off by TOCs (e.g. new stations, changes to route structure)

1.6 How do the passenger numbers for each TOC building block compare with the same data from the previous occasion? Provide proof of changes signed off by TOCs (e.g. growth in a particular route due to increased service frequency). As necessary, amend figures if TOCs provide clear evidence that their alternative is correct and note this.

Stage 2 – creation of sampling plan

Stage 1 provides a list of stations in each TOC building block, with passenger numbers for each. In each building block, the stations are ranked in descending order by passenger numbers and a cumulative percentage produced. The top 25% are assigned to a "very large" subgroup, the next 25% to "large", the next 25% to "medium" and the bottom 25% to "small" (box 1 below).



Each TOC building block has a target sample size and using an assumed number of returns per shift, it is possible to determine how many interviewing shifts are required to generate the required sample. These shifts are produced using a probability proportional to size (pps) approach, meaning that the larger stations are more likely to be selected. This is as expected as more passenger journeys commence at the larger stations (box 2 below).

Where stations are in several TOC building blocks, the number of shifts required can be aggregated to compute the total number of shifts at each station. This is then compared with the sampling plan in the previous wave (box 3 above):

- Where less shifts are required than in the previous wave, remove shifts at random until the required number is left (box 4 above)
- Where additional shifts are required, compare the day of week profile and likely journey purpose profile of the existing shifts with the TOC profile (box 5) and generate additional shifts at the station which are likely to help meet the TOC profile by weekday/weekend and by journey purpose (box 6) – by adding shifts at appropriate days of the week and time of day

This process raises some QA questions, which should be addressed each wave:

2.1 Confirm the allocation of each station in a TOC building block to a station size band – very large, large, medium and small – and store the percentage of total passengers each of the four size bands contributes

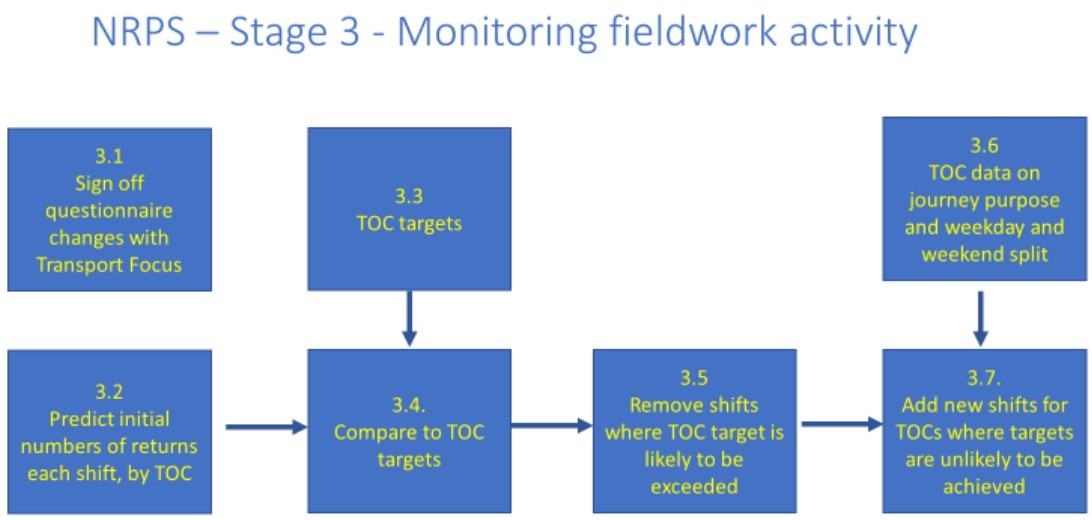
2.2 How does the sampling plan (number of shifts per station across all TOC building blocks) compare with previous wave? Are there any large increases or decreases? If so, provide evidence on why this might be the case (new route introduced, redevelopment of station, relatively new station where numbers are still increasing etc)

2.3 If number of shifts has decreased, confirm which shifts have been removed

2.4 If number of shifts has been increased, confirm the new shifts and specify how days of week and time of day were allocated for each. Confirm TOC data for weekday/weekend split and journey purpose split has been signed off by TOC

Stage 3 – monitoring fieldwork activity

Once the sampling plan has been produced and agreed, NRPS fieldwork takes place. During fieldwork, there are lots of checks that take place to ensure the final sample best meets the objectives.



The research agency undertaking NRPS has a model of how many returns from each TOC are expected from each shift (box 1 above). This model will be based upon:

- The numbers returned on previous waves, if the shift has been undertaken before
- Estimated numbers, based upon the day of week, time of day and size of station being surveyed, for a shift that has not been undertaken before

As shifts are undertaken, the expected numbers are replaced by the numbers achieved and updates made to the likely total for each TOC and each TOC building block (box 2 and 3 above). If estimated numbers exceed targets, it may be possible to cancel relevant shifts (box 4 above); if estimated numbers start to fall below target,

additional shifts (top up shifts) may be required. To maximise efficiency, top up shifts are normally undertaken at the larger stations and on days of the week and times of day that are likely to generate above average returns (boxes 5 and 6 above).

During this process, a number of quality assurance concerns need to be addressed:

3.1 Has the questionnaire had formal sign off from Transport Focus? Has the online version been updated and tested?

3.2 Has the prediction model been initially updated with the details of the new sampling plan?

3.3 Have the TOC and TOC building block targets been signed off by TOCS?

3.4 At regular intervals, have the estimated numbers been replaced with actual returns? If returns are particularly low for individual shifts have these been investigated (e.g. are shifts from a particular interviewer lower than expected, are returns from a particular station lower than expected and why)?

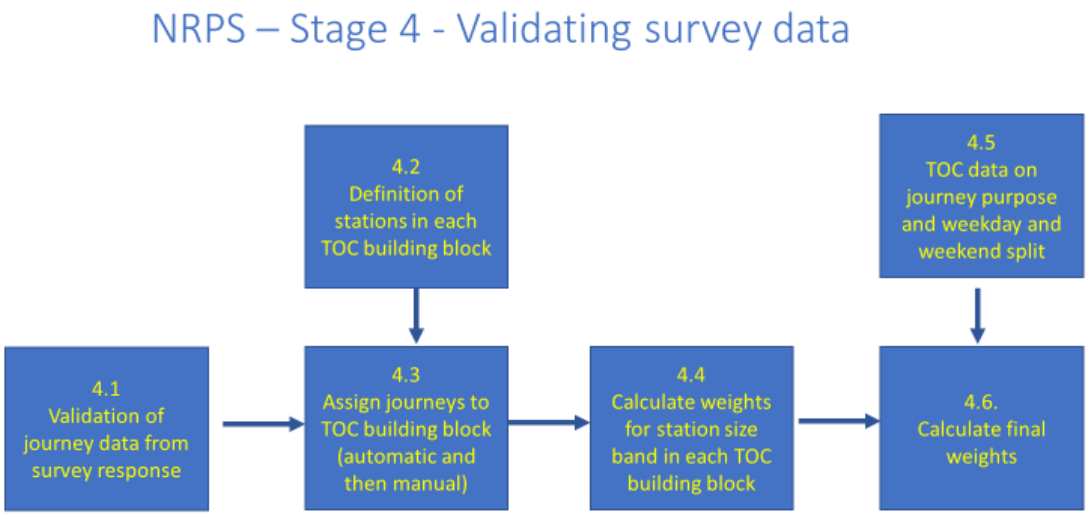
3.5 Where there is a surplus, which shifts have been cancelled and why have those shifts been chosen?

3.6 Have the TOC profiles by journey purpose and weekday/weekend been confirmed?

3.7 Where there is a deficit, which shifts have been added and why have those shifts been chosen?

Stage 4 – validating survey data

As fieldwork takes place, survey responses arrive by post and online. A large number of data validations are applied to the responses before they are admitted to the survey dataset.



Some of the validation rules are fixed, some require updating each wave and so it is important to ensure that the most up to date and agreed rules are used. The current rules will be described in the NRPS Technical Report and the following quality assurance concerns need to be addressed:

4.1 Have the key fields been completed by the respondent (time of journey, TOC used, destination station etc)? If so, does the journey specified exist on the rail timetable for the day of the journey

4.2 Have the definitions of which stations comprise each TOC building block been updated and signed off by the TOC?

4.3 Have the computer rules for automatic assignment of a journey to a TOC building block been updated to reflect any such changes? Have manual assignments been signed off by Transport Focus?

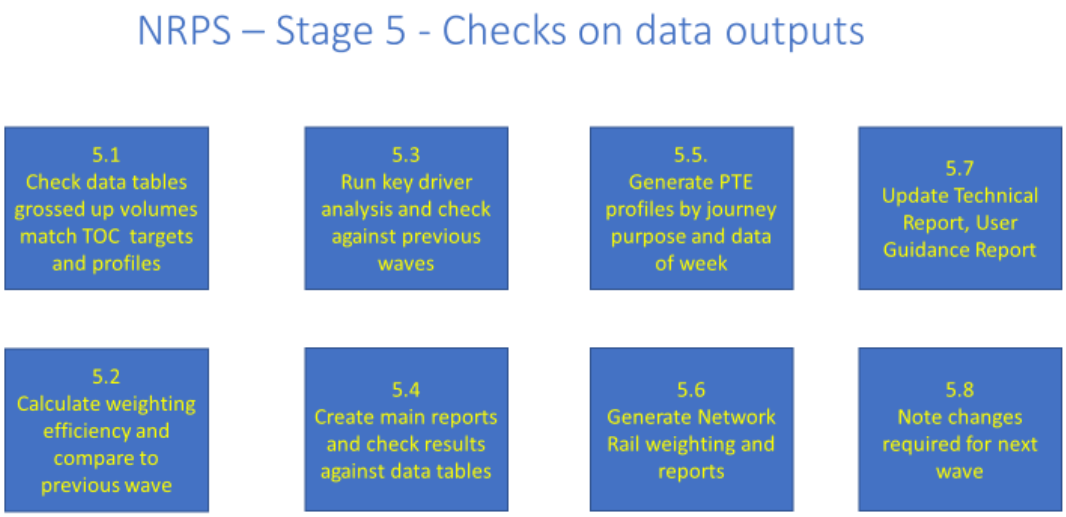
4.4 Are the station size band weights for each TOC building block similar to the previous wave? If not, is there an explanation for this (e.g. amendment to the sampling plan to correct previous under or over representation)

4.5 Have the TOCs updated and signed off total passenger volume numbers and profiles by weekday/weekend and journey purpose? Have any changed markedly from the previous wave and if so, has this been challenged and then agreed?

4.6 What percentage of the final weights are outside the range 0.5 to 2? Do these cluster in specific TOCs or TOC building blocks?

Stage 5 – checks on data outputs

Once fieldwork is complete and all data validation undertaken, a final, weighted dataset is produced. This dataset is used to create a large array of reports and subsidiary reports such as PTE Reports and Network Rail station reports are generated using additional weighting information. Key driver analysis is undertaken to identify the key factors that appear to drive journey satisfaction and dissatisfaction for a range of customer groups.



All these outputs need to be quality assured before they are released to users. The range of checks to achieve this quality assurance include:

5.1 Do the weighted totals in the data tables match the input targets for each TOC by weekday/weekend, journey purpose and station size band within each TOC building block?

5.2 How does the weighting efficiency compare to previous waves? Are there particular TOCs with below average weighting efficiency? If so, what is the cause of this – is it failure to meet day of week, journey purpose or TOC building block targets?

5.3 How does the key driver analysis compare to the previous waves? Are any shifts in either the factors emerging or the coefficients of factors consistent with expected changes (e.g. during a period of declining punctuality, one might expect punctuality to be a bigger driver of journey satisfaction than before)

5.4 Does the data in the key reports match that in the data tables?

5.5 How does the day of week and journey purpose profile for PTE areas compare to that currently used to create the PTE reports? Should the PTE profiles be updated?

5.6 How do the Network Rail targets vary from those used previously?

5.7 Have the specific rules and processes used this wave been updated in the Technical Report and the User Guidance Report?

5.8 Have recommendations been made for updating the sampling plan next wave been recorded and discussed with Transport Focus (e.g. switching some shifts from weekday to weekend for a TOC to reduce day of the week weighting next time)?