NATIONAL PASSENGER SURVEY

Report on an
Audit of the NPS

August 2008

RMA

EXECUTIVE SUMMARY

1.   We are aware of only one error made in generating the W18 data. This was a manual error in the data entry of station size strata weights for new TOCs, which has since been corrected.

2.   Procedural changes have been suggested which we believe will make this error less likely in future.

3.   The Rosetta process has been tested thoroughly and is evidently robust. Also, Rosetta reduces future exposure to human error by automating some of the tabulating and reporting processes.

4.   We believe the corrected W18 data is robust.

5.   Broadly, we believe the real historical data for Waves 1 to 17 to be about as robust as is necessary or practical.

6.   Changes to procedures have been suggested to make checking processes more transparent and effective.

7.   Given the complexity of the rail network, past changes to TOC franchises, and the complexity of the survey structure and data, it is perhaps surprising that the recent checking process revealed so few anomalies in the real historical NPS data.

8.   However, it should be noted that the "virtual historical data" (i.e. past data reanalysed to approximate to a virtual historic result for TOCs which did not exist at that time), will always be less robust than data based on a sample designed specifically to represent a real TOC at that time.

9.   The current approach to sector scores (for Long Distance, London/SE, and Regional) has obvious weaknesses, and the "virtual historical" sector averages are becoming increasingly artificial constructs. We have proposed an approach which will eliminate these issues and generate a much more stable and meaningful set of measures.

Contents   Page

INTRODUCTION

In June 2008, RMA was commissioned by Passenger Focus to conduct an audit of the processes involved in creating the National Passenger Survey (NPS) data and reports, principally re: Wave 18 (Spring 2008) but also re: historical data.

A summary of our conclusions was delivered to Passenger Focus on July 7th, and that was followed by this more detailed report.

COMMENTARY

1.   AUDIT OBJECTIVE

The main objective of the audit was to provide an independent and objective opinion to the Passenger Focus Accounting Officer and NPS Governance Board on whether the controls over the allocation of NPS data to train operating companies (TOCs), and especially any revisions to historical data as a result of changes in the rail franchises, have worked effectively for the Spring 2007 to Spring 2008 waves (Waves 16 to 18). A further objective was to assess whether adequate checking processes were in place for future NPS survey waves. The audit was to be competed by July 2nd, in order to be available for the NPS Governance Board meeting one week later.

The audit scope was to include an independent review of the results of two sets of checking processes already being undertaken by Continental Research which, at the time RMA was commissioned in early June, were expected to be completed by June 9th and 16th respectively. In the event, the results of these checking processes were not available until June 26th and 27th, and the original timetable could not be met.

As a result, it was agreed that the timetable and some other aspects of the terms of reference should be adjusted, and a summary of our conclusions was accordingly delivered on July 7th (the contents of which were broadly as included in the foregoing Executive Summary), followed later by this more detailed report.

2. AUDIT APPROACH

The audit process included a review of a number of NPS reports and other documents, provided by Passenger Focus and Continental, describing various elements of the NPS design for Wave 18 and earlier waves (see Appendix A).

Continental had been carrying out extra manual and electronic checks to see that the W9-17 data shown in the W18 report was the same as that in the W16/17 reports, for TOCs that existed in all three waves and at the national level, and rechecking that the London Midland, East Midlands and CrossCountry data corresponded with that in their virtual TOC reports for previous waves. Comparisons were also made between tabulations produced using the Rosetta system and tabulations produced using the previous Quanvert/Excel method (from identical data). This work generated several large spreadsheets (see Appendix A) which were also reviewed as part of the audit,

Both members of the project team met with senior members of Continental Research to review the processes employed in the execution of the NPS survey. Topics discussed included the following NPS procedures:

*   NPS sampling method (especially re: TOC samples)
*   NPS weighting and analysis methods re: current v. historical data
*   NPS weighting and analysis methods re: real v. virtual TOCs
*   NPS tabulation method (Waves 9-17 Quanvert Excel method)
*   NPS tabulation method (Wave 18 Rosetta method)
*   NPS current checking procedures

The following non-standard elements were also reviewed and discussed:

*   NPS extra manual and electronic checks (on Waves 9 to 18 as above)
*   NPS possible future checking procedures

This was followed by a further examination of the data in the light of our discussions with Continental. Further information or clarification was requested on certain detailed issues, and there were many further telephone conversations and exchanges of emails between the review team, Continental and Passenger Focus.

3. AUDIT REPORT

3.1 Wave 18 Data

Having reviewed the processes and procedures employed in producing, checking and re-checking Wave 18 survey data, and having examined the detailed checking tables themselves, we are aware of only <u>one</u> error affecting NPS Wave 18 survey data.[1]

This was due to a single, manual input error in which station size strata weighting values for the new TOCs introduced for Wave 18 (East Midlands Trains, London Midland, and CrossCountry) were incorrectly inverted: i.e. the weighting factors for: Very Large, Large, Medium and Small stations were applied instead to Small, Medium, Large and Very Large respectively.

This error was originally detected because it resulted in anomalous results for East Midlands Trains.[2] The weights for East Midlands varied significantly by station size, so the impact of the inversion on East Midlands' weighted results was noticeable.

The W18 data has since been rerun with correct weights and we believe the W18 results are now a correct reflection of the intended survey design.

For W19 onwards, Continental have proposed that an extra checking stage (by someone other than the person responsible for entering weights) is applied to the weighting process. We concur that this will reduce the likelihood of such an error.

Also, we believe that the fact that the strata are identified in the database as "1, 2, 3 and 4" made this error possible. Had they been identified as "VL, L, M, S" the error would have been much less likely, and would have been much easier to spot during checking. We recommend that this change is made for future Waves.

On a more general issue: some of the NPS design changes we recommended in the 2006 NPS Review have been implemented, and further enhancements may be possible and desirable in future, but we believe that the Wave 18 NPS design and data are generally robust.

---

[1] I.e. <u>data</u> from Survey Wave 18. Wave 18 <u>reports</u> also contain historical data (discussed later).
[2] This error affected all three of the new W18 TOCs. However, because weighting for the other two new TOCs happened not to vary as much by station size strata, the London Midland results were altered only marginally by the weighting inversion, while CrossCountry results were likely to have been altered by only a fraction of 1%, if at all, and such differences would be undetectable in reports.

3.2   Rosetta

Wave 17 was the first NPS survey wave which employed Rosetta, an established software application for semi-automated reporting of data in tabular or chart form.

The previous procedure was to export data from the Quanvert database into Excel spreadsheets to create tables and charts, and these were then manually pasted (electronically) into the NPS reports.

In contrast, the Rosetta software is fed with a series of parameters which identify the source, destination and format of each table or chart required, and the software then automatically extracts data from the database and sets up tabulations or charts as specified.

A large number of specific checks have been carried out which establish that, using the same data, and given correct input of parameters, the results produced by Rosetta are effectively identical to those produced using the previous method.[3]

Some manual input processes will remain, but they will be far fewer with Rosetta than with the previous process (for example, *ceteris paribus*, all the tables from W18 could be repopulated with new data from W19 with little, if any, further manual input).

Therefore, the Rosetta method is reliable, and more robust than the previous method.

---

[3] The only difference relates to the number of decimal points used in calculations: Quanvert calculates to 3 dp, while Rosetta calculates to 2 dp, which may affect rounding. For example, any Quanvert value between 50.395 to 50.399 would be shown as 50.40 by Rosetta (i.e. overstating the value by less than 0.01%); and rounded to whole numbers, all would appear as 50. Taking a more extreme case, Quanvert values between 50.495 to 50.499 would be shown as 50.50 by Rosetta (again overstating the value by less than 0.01%); but if these values are then rounded to whole numbers, Quanvert would show 50 and Rosetta 51. However, even this difference (which is likely to be statistically insignificant for any NPS TOC results) can occur only on the small percentage of all results ending around .500, and only when comparing 2 dp source data with 3dp source data, which is unlikely to occur in future.

3.3   Historical Data for Real TOCs

Broadly, we believe that the historical data for Waves 1 to 17 are probably about as robust as is necessary or practical, where they relate to "real" TOCs (i.e. TOCs which existed at the time).

Given the complexity of the rail network, changes to TOC franchises over the years, and the consequent complexity of the survey structure and data, it is not surprising that the recent checking process has revealed some apparent errors or anomalies in historical data relating to TOC definitions (these are identified in Continental's NPS Report Review)[4]. Indeed, it is perhaps surprising that so few were found.

In future, improved checking procedures, and the use of Rosetta (which will reduce manual input), will reduce the likelihood of errors.

---

[4] For example, Continental reported that two Regional TOC Reports for Wave 14 inadvertently included Heathrow Express in their comparative regional data. Heathrow Express is classified as a regional TOC but was not included in sector figures generally for Wave 14. The other regional TOC reports for Wave 14 correctly excluded Heathrow Express.

3.4   Current Data for New TOCs

The NPS sample design is based on the principle of designing a sample for each existing TOC, structured to be representative of that TOC. This allows sample sizes to be set for each TOC, and by aggregating the TOC samples and weighting the results correctly, an overall weighted result is achieved which is representative both nationally and for each TOC individually.[5]

One element of the sample design at TOC level is that total journey volumes need to be estimated for each station in each TOC's sample (and where stations are shared with other TOCs, journey numbers need to be apportioned).

If a franchise change merely involves a change of livery then the sample structure can remain the same.

If the change involves two TOC networks merging into one larger TOC, the sample structure can remain largely the same, but the sample base size will need adjustment.

Where TOC changes involve boundary, route or frequency changes (rather than merely a change of livery or a merger), then the solution can be rather more complex. Early notification of TOC boundary changes would be helpful, but this may not always be possible. Even so, this does not create insoluble problems.

However, it should be noted that volume estimates available for sampling purposes re: a new TOC may be less reliable in the first year than in subsequent years, when the actual journey numbers are available.

Any TOC changes will of course require careful checking and amendment to the sample design and reporting framework, but none of the above scenarios creates serious problems for the researcher.

---

[5] In some cases the building blocks employed are parts of TOCs, rather than whole TOCs, which is useful e.g. where a new TOC consists of former parts of other TOCs.

3.5   Virtual Historical Data for New TOCs

The creation for a new TOC of "virtual historical data" (i.e. manipulating past NPS data to create an approximation of what past performance was on routes taken over) has a logical purpose and value. It could, for example, be helpful in setting performance goals for a new TOC.

However, it must be recognized that such "virtual historical data" will be less reliable than would "real" data, particularly if there have been TOC route or boundary changes over the period. Attempting to show past results for TOCs which did not then exist, based on aggregating past data from parts of samples designed for former TOCs, can never be as robust statistically as real historical data for a real TOC (i.e. where data is based on a sample designed specifically to represent that TOC at that time).

Also, it appears that errors may be more likely to occur when dealing with virtual historical data because of the greater complexity of the exercise. For example, in order to create virtual historical trend data for London Midland, a complicated filter had to be applied to earlier waves. In simplified outline form it was:

*   Wave 18 & TOC = London Midland or
*   Wave 17 & (TOC building blocks = Silverlink County or
        relevant Central building block) or
*   Wave 16 and earlier & (TOC building block = Silverlink County or
        Central records assigned to this part of the franchise)

A misplaced bracket resulted in a logic error, which has now been corrected.

Wave 16 and earlier were particularly complicated because individual journey records were selected from the Central franchise data rather than using a building block approach (as the TOC was not defined until just prior to W17). Filtering problems can be minimised if pre-designated building blocks are used rather than having to assign individual records to a virtual TOC, but this requires earlier notification of TOC boundaries etc., which may not always be possible.

However, better checking procedures (already planned) will reduce the chances of associated errors.

3.6   Sector Data

The sectors are currently defined by aggregating TOCs, i.e. all TOCs are classified as being in one of three categories: London & SE, Long Distance, or Regional. The averages for each category are then reported as "sector averages" for London & SE, Long Distance, and Regional. This situation arose from certain TOC franchise contract issues. However, a weakness of this approach is that it ignores the fact that many TOCs have a mix of route types (increasingly so in recent years).

A second weakness of the approach is that whenever new TOCs are created, or their boundaries change, the sector averages (as defined) will also change. Comparisons of current sector scores with past sector scores will then become invalid. To overcome this, historical sector scores are recalculated using the latest sector definitions.

These "virtual historical" sector averages are becoming increasingly artificial constructs.

Serious thought should be given to devising a more realistic and stable measure. The old "sector scores" could easily continue to run in parallel as long as they are still required for contractual reasons, but should be phased out as soon as possible.

3.7   Proposed Sector Data

A new, objective and stable sector measure could be derived very easily from the NPS. For example:

Actual Long Distance
It would be relatively simple (probably within existing procedures) to identify all NPS journeys which had a scheduled duration of, say, 90 minutes or more (alternatively, it could be defined by distance travelled). This would allow the NPS to provide sector scores for <u>actual</u> "Long Distance Journeys", undiluted by other traffic, both in total and for any TOC which operated any.

Actual London & SE Journeys
It would be relatively simple also to identify all journeys which started <u>and</u> ended within a defined, agreed area of the SE of England. This would allow the NPS to provide actual scores for the "London & SE" Sector, undiluted by other traffic.

Actual Regional
All NPS journeys not in the Long Distance or London & SE sectors would be defined as "Regional".

The advantages of this approach would be:

•   much more precise definitions
•   sector definitions will not change over time
•   obviates the need to create virtual historical scores
•   all TOCs can compare like with like

The cost would be negligible in the context of the overall NPS budget.

Similarly, useful sector-type average scores could easily be added re:

Weekday Morning Peak (e.g. journeys departing Mon-Fri 07:00 - 09:59)[6]
Weekday Evening Peak (e.g. journeys departing Mon-Fri 16:00 - 19:59)
Off-Peak (the rest)

Once again, the cost would be negligible in the context of the overall NPS budget, but, like the proposed sector scores above, it would allow comparisons to be made between TOCs, which would be unaffected by the variations in mix between TOCs.

---

[6] These are the peak am and pm departure times, based on the analysis shown in Section 3.4 of "Findings of a review of the NPS", PF/RMA February 2006.

3.8 Checking Procedures

Continental already had various quality control checking procedures in place prior to this audit. As an MRS Company Partner, Continental adheres to the MRS Code of Conduct. Also, Continental adheres to IQCS benchmark standards on fieldwork. Various standard checks are applied to the NPS, and some processes are designed specifically for checking the NPS, for example: NPS journeys are checked against the electronic timetable, and departure time and journey are checked for feasibility.

We recommend also that the extra checking procedures discussed with Continental (and mostly suggested by them) are added to the standard NPS procedure. These include, for example, a check on data entry for weighting to be carried out by a second individual (not by the individual who entered the weights).

However, our main recommendation in respect of checking procedures is that all the existing and proposed new checking procedures that Continental Research applies to the NPS should be fully documented, in order to make the process more transparent and better understood. Consideration should also be given to adopting ISO 20252 standards (the replacement for MRQSA/BS 7911).

APPENDICES

Appendix A: Items Reviewed

Appendix B: Project Team

Appendix A   Items Reviewed

Items reviewed included, but were not limited to, the following:

Documents

NPS Questionnaire Autumn 2007 (W17) - Master Copy v7
NPS Questionnaire Spring 2008 (W18) - Master Copy v6
NPS Report Autumn 2007, PF/CR, 2008
NPS Report Review, CR June 2008
NPS Technical Survey Overview Wave 14, CR 2006
NPS Technical Survey Overview Waves 17-18, CR 2008
NPS User Guidance, CR, 2007
Findings of a review of the NPS, PF/RMA February 2006

Spreadsheets

Wave 16 Summary report check vs Rosetta.xls
Wave 16 Summary report check vs Wave 18.xls
Wave 16 TOC Central vs Rosetta.xls
Wave 16 TOC gner vs Rosetta.xls
Wave 16 TOC northern vs Rosetta.xls
Wave 16 TOC one vs Rosetta.xls
Wave 16 TOC southern vs Rosetta.xls
Wave 16 TOC vwc vs Rosetta.xls
Wave 17 Summary report check vs Wave 18.xls
Wave 18 Summary report check vs Quanvert database.xls

KEY:

CR = Continental Research
PF = Passenger Focus (Rail Passengers Council)
RMA = Roberts-Miller Associates

Appendix B   Project Team

The project team consisted of Richard Roberts-Miller and Gary Bennett, both of whom were also members of the project team that carried out the 2006 NPS Review for Passenger Focus ("Findings of a review of the NPS", PF February 2006):

Richard Roberts-Miller, Roberts-Miller Associates

Richard established Roberts-Miller Associates (RMA) in 1989 as a market research, marketing and management consultancy. Initially specialising in travel, tourism and transport, the range of public and private sector clients has grown to include those in other complex service industry fields such as finance, education, health, and market research. The work has included marketing, advertising and branding projects, strategic management reviews, takeover and merger projects and investment project evaluation studies, as well as many market research surveys (quantitative and qualitative, consumer and B2B, domestic and international), customer satisfaction surveys, research design projects and technical evaluation studies.

Prior to RMA, Richard Roberts-Miller worked at the British Market Research Bureau (as a research executive), Times Newspapers (senior research executive), The Thomson Organisation (group market research manager), Thomson Travel (research & planning manager), Thomson Holidays (marketing controller), Thomson Vacations USA (President & CEO), and Thomson Travel Inc. North America (President & CEO). He has delivered papers on a range of market research topics to MRS, ESOMAR, the Marketing Society, ADMAP and others. He served on the Jamaican Government Tourism Advisory Council, and received the Jamaican Blue Mountain Award for services to tourism. He is a BSc Economics graduate of Southampton University, a Fellow of the Market Research Society, a Fellow of the Royal Geographical Society and a Fellow of the Tourism Society.

Gary Bennett, Logit Research

Gary Bennett established Logit Research as an independent marketing research consultancy in March 2004. Logit provides a full-service consultancy to market research agencies, field & tab and data analysis companies and also works with end users of research, both directly and in conjunction with partners. Logit offers a portfolio of tools and services primarily focused on adding value to research through the use of advanced multivariate statistical and modelling techniques. The work ranges from the design and analysis of choice and stimulus exercises needed to build models of behaviour from research data, to providing provide guidance on the statistical confidence around the resulting data and models. Logit Clients have included major industrial and financial organisations as well as very many research agencies and consultancies.

Gary Bennett has 16 years experience in the UK market research industry, including half a decade in a director role. Prior to establishing Logit, Gary worked at MORI, RS Consulting, Benchmark Research and as a client-side market researcher/analyst at London Transport (now TfL). Gary is a Full Member of the Market Research Society, a Fellow of the Royal Statistical Society and an Associate Member of the American Marketing Association.