

Task 1.6 Public Transport Survey Processing (example of Auckland)

Bus Survey Processing

Data Acceptance

Questionnaires

The provided questionnaire data spans 1076 unique bus trips. Adding to this number the 28 bus trips where no questionnaire data was collected¹, and 6 ‘two/four-part trips’² gives 1112 trips, which is the total number of bus trips performed by survey staff, as recorded by Gravitas. Therefore we have questionnaires from all surveyed bus services.

Questionnaire data was received from Gravitas (survey consultants) in the form of Microsoft Access database. The total number of questionnaires received is 14,985.

For each questionnaire the MIC requirements are that there be:

- no missing data on Key Items of information (Q1-6 on survey: O/D purpose, address & mode of connecting transport);
- no more than 2 missing responses for all other items (questions).

12,473 questionnaires meet the first bullet point of the Minimum Information Content (MIC) requirements and 12,319 satisfy both bullet points.

The final MIC test is that across the entire sample of questionnaires, there shall be no more than 5% of missing responses for each other item. Of the 12,319 passing questionnaires, all questions have no more than 2% missing responses (Table 6).

The final data set comprised 12,319 good questionnaires, in excess of the survey target of 10,000 questionnaires.

Table 1 – Missing responses for non-key questions

	MIC Passed 'Good'	
Questionnaires	12319	
	Missing Responses	Percentage
Q7	45	0.4%
Q8	184	1.5%
Q9	167	1.4%
Q10	167	1.4%
Q11	139	1.1%
Q12	100	0.8%
Q13	136	1.1%
Q14	36	0.3%
Q15	224	1.8%
Total	1198	1.1%

¹ Due to either no passengers boarding, no surveys being handed out, or no surveys being returned
² Where most passengers stayed on the bus as it started a new ‘trip’, effectively treating it as a single bus route. There were 5 ‘two-part’ trips and 1 ‘four-part’ trip.

- *Survey Control Data*

In addition to managing questionnaires, survey staff collected data on boarding and alighting passengers on each surveyed bus. This data lists, for each bus trip and for each stop:

- number of boarding and alighting adults and <15's;
- number of refused questionnaires, handed out questionnaires and other information;
- the date, time and weather conditions.

The data are complete and in a format suitable for use in expanding questionnaire responses up to all passengers on each bus trip.

Range and Logic Checks

Checks were performed on the data to ensure sensibility of the received data. For most questions, responses were coded in defined manner – therefore not requiring range checking.

Expansion of Questionnaires to Surveyed Buses

This process expands the survey questionnaires to represent the total passengers on the surveyed buses. The process is in 5 separate stages, outlined in Table 2.

Table 2 – Bus Expansion Factor Summary (Expansion to Surveyed Buses)

Expansion Factor	Description
EF01Adults	By bus route and bus stop, expands the collected questionnaires to the number of in-scope (aged > 15) boarding passengers.
EF02Capping	The maximum value of EF01 that was permitted was 10. Larger values were factored down to be equal to 10. This affected a few cases only.
EF03Accomp	<p>Respondents were also asked how many children (between 4 and 15 year old) were travelling with them. This factor expands the questionnaires to represent the trips of accompanying children.</p> <p>Example 1: For an adult travelling with a two children, EF03 would be 3 (1 adult + 2 children).</p> <p>Example 2: If a respondent was travelling with an adult and three children, the three children are “shared” between the two adults, so they would have 1.5 children each, so the EF03 would be 2.5 (1 adult + 1.5 children).</p>
EF04Unaccomp	<p>The remaining unsurveyed trips related to unaccompanied children (age <15). These trips were identified as the difference between</p> <ul style="list-style-type: none"> ■ the total number of accompanied children boarding at a stop, and ■ the total number of children counted boarding by survey staff. <p>We apply an additional expansion factor to account for these trips.</p>
EF05TimeBias	<p>When comparing blank/poorly filled in questionnaires to accepted questionnaires, a bias was identified which resulted in short trips being under-represented (as a result of passengers not having enough time to fill in the questionnaires). We compared the trip time characteristics of returned completed and returned blank questionnaires (for which we nevertheless had bus stop data) to determine to what degree shorter trips were missing in the database, and a set of travel-time based factors were developed. These factors have the effect of giving more weight to questionnaires filled in on shorter trips, in order to restore the balance of short/long trips in the final data. Table 3 shows these factors. As can be seen, questionnaires filled in on trips of less than 5 minutes duration were factored up by the greatest amount, reflecting the difficulty respondents had completing a survey in this time.</p>

Table 3 – Time Bias Factors

Time Interval	Time Bias Factor
0-5 minutes	1.507
6-10 minutes	1.085
11-15 minutes	0.991
16-20 minutes	0.971
> 20 minutes	0.925

The representative population for each level of expansion is shown in Table 5.

Expansion to Full Weekday using ETM data

- *Pre-processing of ETM data*

ETM patronage data was received from six operators: Stagecoach, Birkenhead, Howick& Eastern, Pavlovich, Ritchies and Bayes. As Bayes only operated school buses, their data was not required.

Most data was provided in a simple spreadsheet format, with patronage provided for each combination of hour of day, ticket type and 'route'. Stagecoach also included direction of travel. "Route" was provided as a *RouteUID* a unique identifier used by the operators which represent variations on the publicly viewed route numbers, such as direction of travel, deviations, or early or extended termination locations. Birkenhead provided data indexed by a different route ID number system, which is related to, but different from RouteUID.

Mapping the various ID numbers to routes, and therefore to corridors for use in expansion was not a clean task. Expansion factors EF08 and EF09 (see Table 4) account for the data that was unable to be mapped to a precise corridor. 'Missed' counts of this nature were incorporated into the expansion process as locally as possible using the available information. Only count data for which there was no identifiable route, corridor or region was used for EF09, which is a uniform expansion factor applied to all questionnaires.

All data was rationalised into two directions – *in* and *out*. Patronage counts for which there was no provided direction were split 50/50 between the two directions. Anticlockwise/clockwise directions were assigned in/out according to the convention used by the survey agency, Gravitass.

The expansion to ETM patronage counts was attempted on a per-corridor, per-time period, per-direction basis. Where the number of questionnaires was deemed insufficient (generally less than 10), and/or where very large expansion factors would have resulted (generally greater than 15 or 20), corridors were merged, based on an acceptable grouping. This was accomplished by defining groups of similarly aligned corridors, which could be merged in particular instances where necessary. In some cases the School and Interpeak time periods were merged in addition to/instead of merging corridors. In all cases, merging was performed minimally, and only where deemed necessary.

Table 4 – Bus Expansion Factor Summary (Expansion to Count Data)

Expansion Factor	Description
EF10 TimePeriod	<p>Expansion by Time Period: For each combination of time period, (merged) corridor and direction, this factor expands the passengers on the surveyed buses (as represented by the expanded questionnaires - Table 2) to represent the full bus service, as given by the total ETM passenger count. This expansion is by time period³, corridor and direction.</p> <p>This is the primary expansion factor to ETM Counts. EF06-EF09, which follow, are smaller in magnitude and only account for small volumes of passengers.</p>
EF06Missed	<p>Missed Corridor/Time Periods/Directions: In some corridor/time period/direction combinations, no questionnaires had been obtained in the surveys but passengers had been counted in the ETM data. This factor spreads these additional passenger counts across questionnaires in all time periods for the given direction and corridor.</p>
EF07Offpeak	<p>Offpeak Passengers: Passenger counts for the offpeak period⁴ which was not surveyed were also spread amongst questionnaires in the specified corridor.</p>
EF08 RegionUnall	<p>Region Unallocated Counts: ETM passenger counts for which a specific corridor could not be determined, but for which the region (North, South, East, West, Central or Crosstown) was known were spread amongst all questionnaires in that region.</p>
EF09 OverallUnall	<p>Overall Unallocated Counts: ETM passenger counts for which no geographical information could be determined at all were spread over all questionnaires in the database.</p>

The representative population for each level of expansion is shown in Table 5.

³ Surveyed Time Periods: AM Peak 7-9am, Interpeak 9-3pm, School 3-4pm, PM Peak 4-6pm. See Section **Ошибка! Источник ссылки не найден., Ошибка! Источник ссылки не найден..**

⁴ Unsurveyed Time Period: Offpeak 6pm-7am.

Expansion Summary

Table 5 summarises the magnitude of Bus PTIS Questionnaire expansion factors.

Table 5 Representative Populations for Bus Expansion Stages

Expansion Factor	Expansion Step	Mean Expansion Factor	Population
-	Questionnaires	-	12,199
EF01	Adults	1.66	20,283
EF02	Capping EF01	0.99	20,153
EF03	Accompanied Children	1.02	20,565
EF04	Unaccompanied Children	1.13	23,211
EF05	Time Bias Factor	1.01	23,450
EF10	Pax with known time period, corridor, direction	6.31	148,071
EF06	Missed Pax (in corridor & time period)	1.00	148,698
EF07	Offpeak passengers	1.04	155,077
EF08	Missed Pax (in region)	1.01	157,181
EF09	Missed Pax (unknown region)	1.05	165,419
EF11	Trip reversal	N/A	N/A
Overall		13.56	165,419

Ferry Survey Processing

Data Acceptance

■ Questionnaires

The provided *questionnaire* data span 107 unique ferry services. This matches the total number of ferry trips performed by survey staff, as recorded by Gravitas. Therefore we have questionnaires from all surveyed ferry services.

Questionnaire data were received from the survey consultants (Gravitas) in the form of Microsoft Access database. The total number of ferry questionnaires received was 2,679.

For each individual questionnaire, the MIC requirements were that there be:

- no missing data on Key Items of information (Q1-6 on survey: O/D purpose, address & mode of connecting transport);
- no more than 2 missing responses for all other items (questions).

2,260 questionnaires meet the first bullet point of the Minimum Information Content (MIC) requirements specified above and 2,235 satisfy both bullet points. The second point has been removed as a MIC requirement.

The third MIC test is that there should be no more than 5% of missing responses for each other item (i.e. none Key Item) across the entire sample of questionnaires.

Of the 2,260 questionnaires which met the first two MIC requirements, all questions have no more than 2% missing responses (Table 6), except Question 8: "What time did you, or will you, make this return trip today?" for which 11.6% of responses were left blank. Although this would result in larger expansion factors for reversed trips, it was deemed to be acceptable.

The final data set comprised 2,260 good questionnaires, in excess of the survey target of 2,000 questionnaires.

Table 6 – Missing responses for non-key questions

	MIC Passed 'Good'	
Questionnaires	2,260	
	Missing Responses	Percentage
Q7	43	1.9%
Q8	262	11.6%
Q9	15	0.7%
Q10	17	0.8%
Q11	23	1.0%
Q12	23	1.0%
Q13	27	1.2%
Q14	17	0.8%
Q15	34	1.5%
Q16	12	0.5%
Q17	60	2.7%
Total	533	2.1%

■ **Survey Control Data**

In addition to managing questionnaires, survey staff collected data on boarding passengers on each surveyed ferry. These data list, for each ferry service and for each stop:

- number of boarding adults and <15's;
- number of refused questionnaires, handed out questionnaires and other information;
- the time.

The dataset is as required.

Range and Logic Checks

Checks were performed on the data to ensure sensibility of the received data. For most questions, responses were coded in defined manner – therefore not requiring range checking.

Expansion of Questionnaires to Surveyed Ferries

This process expands the survey questionnaires to represent the total passengers on the surveyed ferries. The process is in 3 separate stages, as outlined in Table 7. The expansion factors have the same meaning as correspondingly named factors in Table 2 for buses.

Table 7 – Ferry Expansion Factor Summary (Expansion to Surveyed Ferries)

Expansion Factor	Description
EF01Adults	By ferry route and terminal, expands the collected questionnaires to the number of in-scope (aged > 15) boarding passengers.
EF03Accomp	<p>Respondents were also asked how many children (between 4 and 15 year old) were travelling with them. This factor expands the questionnaires to represent the trips of accompanying children.</p> <p>Example 1: For an adult travelling with a two children, EF03 would be 3 (1 adult + 2 children).</p> <p>Example 2: If a respondent was travelling with an adult and three children, the three children are “shared” between the two adults, so they would have 1.5 children each, so the EF03 would be 2.5 (1 adult + 1.5 children).</p>
EF04Unaccomp	<p>The remaining unsurveyed trips related to unaccompanied children (age <15). These trips were identified as the difference between</p> <ul style="list-style-type: none"> ■ the total number of accompanied children boarding at a stop, and ■ the total number of children counted boarding by survey staff. <p>We apply an additional expansion factor to account for these trips.</p>

Note: EF02Capping and EF05TimeBiasFactor are not applicable for the Ferry survey.

Expansion to Full Weekday using ETM Data and Trip Reversal

- *Pre-processing of patronage data*

ETM patronage data was received from ferry operators Fullers and Belaire for all surveyed routes except for GulfHarbour and PineHarbour routes, for the 5 week period, 24th July 2006 – 25th August 2006. Additional data was received for a week during May 2006, but the larger dataset was selected for use as it provides a better estimate of the daily average volumes, and the data is closer to the actual dates the survey was undertaken.

The methodology of attempting to sample every ferry service across the day⁵ meant that there was less of a reliance on the ETM data to ‘fill the gaps’ – 105 of the 115 scheduled services were surveyed. The number of boarding passengers was recorded for each ferry service surveyed.

These onboard patronage counts were used for expansion on surveyed ferries, in preference to operator supplied count data. For non-surveyed ferries the ETM data was used for expansion. Due to the survey data and ETM being collected at a similar time of the year, no seasonal adjustment was required when combining survey counts and ETM counts.

In the case where neither form of counts were available (as was the case for two services in the PM Peak on the HalfMoonBay route), a technique was used which allocated counts to the two missing services based on the profile of PM peak demand on other ferry routes.

- *Expansion to patronage counts*

The expansion factors used for ferry expansion were similar to those used for the bus expansion, and are named in Table 8 accordingly. Fewer expansion factors were used for ferry expansion, as the quality of the data did not necessitate them.

⁵ From 6:30am to 7:00pm.

Table 8 – Ferry Expansion Factor Summary (Expansion to Count Data)

Expansion Factor	Description
EF06Missed	Missed Timetabled Ferries: On some ferry routes, not all timetabled services were surveyed. Patronage data for these ferry services was grouped together with an adjacently timetabled service on the same route, which had been surveyed. This affected only a small number of services.
EF07Offpeak	Offpeak Passengers: Ferries were surveyed between 6:30am and 7:00pm. Passenger counts outside these times were spread amongst all surveyed ferries on that route.
EF11Reverse	Reversed Questionnaires: Only outbound ferries were surveyed. Questionnaires were reversed to represent the return (inbound) trips made by respondents. Not all questionnaires were able to be reversed, so an additional factor was applied to successfully reversed questionnaires to expand to the total outbound patronage count. This factor was applied on an hourly basis, based on the surveyed trip time. The resulting inbound trip profile provided a sufficient fit to the ETM count profile, so additional profile adjustment was not required. Also see <input type="checkbox"/> Questionnaire Trip Reversal, ниже.

Note: EF08RegionUnall, EF09OverallUnall and EF10TimePeriod are not applicable for the Ferry survey.

■ *Questionnaire Trip Reversal*

For the ferry (and also train) survey, questionnaires were collected only on outbound trips. *Trip reversal* is used to gain a full representation of passengers travelling in both directions. The first step of this process is to ask respondents if they are planning to make (or have already made) a trip in the opposite direction that day, and the time of that trip.

For each respondent that provides information on their return trip⁶, a second questionnaire is generated, which inherits the majority of question responses from the original questionnaire. Questions relating to origin and destination are interchanged to reflect the reverse trip (destination to origin). Categorisation based on travel time is adjusted to reflect the stated return trip departure time. Basically, it is assumed that the respondent effectively retraces their steps, using the same transport modes for access and egress and having the same trip duration. Generated questionnaires also inherit the expansion factors (1-10) of the original questionnaires.

Clearly not all questionnaires are able to be reversed, either due to respondents not satisfactorily completing this question, or stating they were not making a return trip. To reflect this, an additional factor is applied to generated questionnaires, *EF11Reverse*. This factor expands the reversible questionnaires to cover the same population size as all the original questionnaires:

$$\text{Average (mean) of } EF11 = \frac{\prod_{\text{all original questionnaires}} EF1 \dots EF10}{\prod_{\text{reversible questionnaires}} EF1 \dots EF10}$$

⁶“Return trip” in the context of trip reversal to refer to the trip made in the non-surveyed direction, regardless of which trip is made first.

This factor was calculated taking into consideration the departure time of the surveyed trip. It was found that the proportion of reversible questionnaires was greater before 9am, with lower proportions of questionnaires having return trip information in the afternoon and PM peak (and hence larger expansion factors used for these hours). It is thought that respondents experienced some difficulty with the wording of the question, particularly the term *return trip*⁶ when used in reference to a trip made earlier in the day.

It was important to ensure that the reversible questionnaires were a non-biased sample of the total set of survey responses, lest the generated questionnaires give a skewed account of inbound passengers' characteristics. Bias checks were performed on the reversible questionnaires for key output fields. No significant bias was identified.

Expansion Summary

Table 9 summarises the magnitude of Ferry PTIS Questionnaire expansion factors. Note that there are effectively two sets of questionnaires that were expanded – the original survey questionnaires (for outbound trips), and the questionnaires generated from the responses on *reversible* questionnaires to represent the return trips. Both sets are expanded in the same manner, except for EF11, which only relates to the reversed questionnaires.

Table 9 – Representative Populations for Ferry Expansion Stages

Expansion Factor	Expansion Step	Outbound		Inbound (Reversed)	
		Mean Expansion Factor	Population	Mean Expansion Factor	Population
-	Questionnaires	-	2,260	-	1,005
EF01	Adults	2.05	4,640	2.00	2,011
EF02	Capping EF01	-	-	-	-
EF03	Accompanied Children	1.03	4,768	1.02	2,061
EF04	Unaccompanied Children	1.04	4,937	1.04	2,138
EF05	Time Bias Factor	-	-	-	-
EF10	Expansion by Time Period	-	-	-	-
EF06	Missed Pax (in corridor & time period)	1.06	5,220	1.06	2,260
EF07	Offpeak passengers	1.16	6,043	1.16	2,632
EF08	Missed Pax (in region)	-	-	-	-
EF09	Missed Pax (unknown region)	-	-	-	-
EF11	Trip reversal	-	-	2.30	6,043
Overall		2.67	6,043	6.01	6,043
Grand Total		3.70	12,086		

Train Survey Processing

Data Acceptance

■ Questionnaires

The provided questionnaire data span 78 unique train services. This matches the total number of train trips performed by survey staff, as recorded by Gravitas. Therefore we have questionnaires from all surveyed train services.

Questionnaire data were received from the survey consultants (Gravitas) in the form of a Microsoft Access database. The total number of train questionnaires received was 5,475.

For each questionnaire the MIC requirements are that there be:

- no missing data on Key Items of information (Q1-4, 6 & 8 on survey: O/D purpose, address & mode of connecting transport);
- no more than 2 missing responses for all other items (questions).

4,795 questionnaires meet the first bullet point of the Minimum Information Content (MIC) requirements specified above and 4,723 satisfy both bullet points. The second point has been removed as a MIC requirement.

The third MIC test is that, there should be no more than 5% of missing responses for each other item across the entire sample of questionnaires. Of the 4,795 questionnaires meeting the first two MIC requirements, all questions have no more than 3% missing responses, thereby meeting this criterion (Table 6), with the exception of Question 10: “What time did you, or will you, make this return trip today?” for which 6.5% of responses were blank. This question is used solely for trip reversal, and is not present in the final PTIS output. With this in mind, the dataset was deemed acceptable against this MIC test.

Table 10 – Missing responses for non-key questions

	MIC Passed 'Good'	
Questionnaires	4795	
	Missing Responses	Percentage
Q5	14	0.3%
Q7	15	0.3%
Q9	88	1.8%
Q10	310	6.5%
Q11	53	1.1%
Q12	52	1.1%
Q13	66	1.4%
Q14	66	1.4%
Q15	56	1.2%
Q16	46	1.0%
Q17	62	1.3%
Q18	37	0.8%
Q19	137	2.9%
Total	1002	1.6%

The final data set comprised 4,795 good questionnaires, in excess of the survey target of 2,500 questionnaires.

■ **Survey Control Data**

In addition to managing questionnaires, survey staff collected data on boarding passengers on each surveyed train. These data list for each train service and for each station:

- number of boarding adults and <15's;
- number of refused questionnaires, handed out questionnaires and other information;
- the departure time (for most stations – data was sufficiently dense to allow interpolation of missing points);
- a headcount of passengers on board in each carriage (for some stations only).

This dataset is sufficiently complete and in a format suitable for use in expanding questionnaire responses up to all passengers on each train service.

Range and Logic Checks

Checks were performed on the data to ensure sensibility of the received data. For most questions, responses were coded in defined manner – therefore not requiring range checking. Summaries of the distribution of question responses are given in the separate Survey Data Report.

Expansion of Questionnaires to Surveyed Trains

This process expands the survey questionnaires to represent the total passengers on the surveyed trains. The process is in 3 separate stages, as outlined in Table 11. The expansion factors have the same meaning as correspondingly named factors in Table 2 for buses.

Table 11 – Train Expansion Factor Summary (Expansion to Surveyed Trains)

Expansion Factor	Description
EF01Adults	By train and station, expands the collected questionnaires to the number of in-scope (aged > 15) boarding passengers. Passenger counts at stations for which no good questionnaires exist were aggregated with the nearest stations that did have valid questionnaires returned.
EF03Accomp	Respondents were also asked how many children (between 4 and 15 year old) were travelling with them. This factor expands the questionnaires to represent the trips of accompanying children.
EF04Unaccomp	The remaining unsurveyed trips related to unaccompanied children (aged between 4 and 15). These trips were identified as the difference between <ul style="list-style-type: none"> ■ the total number of accompanied children boarding at a stop, and ■ the total number of children counted boarding by survey staff. We apply an additional expansion factor to account for these trips.

Note: EF02Capping and EF05TimeBiasFactor are not applicable for the train survey.

Expansion to Full Weekday using ETM Data and Trip Reversal

Table 11 indicated the expansion factors used to represent all surveyed trains. Due to the train sampling covering all timetabled trains in the survey time period, there were no ‘missed’ trains. This indicates that the data, as expanded in Table 11 represents *all passengers* in the surveyed time period on all outbound trains. As a result, it was decided to use the ETM data only for expansion into the offpeak (unsurveyed) time period. Table 12 includes the expansion factor which accomplishes this, and an expansion factor related to trip reversal.

As a result of the complete survey coverage, and a well-formed and complete ETM count database, expansion factors EF06Missed, EF08RegionUnall and EF09OverallUnall were not required for train questionnaire expansion.

Table 12 – Train Expansion Factor Summary (Expansion to Count Data)

Expansion Factor	Description
EF07Offpeak	Offpeak Passengers: Trains departing between 6:30am and 7:00pm were surveyed. ETM data was used to calculate the ratio of offpeak passengers to survey period passengers for each train line. This ratio was applied to all surveyed trains on that line.
EF11Reverse	Reversed Questionnaires: Only outbound trains were surveyed. Questionnaires were reversed to represent the return (inbound) trips made by respondents. An additional factor was applied to successfully reversed questionnaires to expand to the total outbound trip count. This factor was applied on an hourly basis, in the same fashion as for the ferry survey ().

Note: EF06Missed, EF08RegionUnall, EF09OverallUnall and EF10TimePeriod are not applicable for the Train survey.

- *Questionnaire Trip Reversal*

The method of trip reversal is explained with regard to the ferry questionnaires.

Expansion Summary

Table 13 summarises the magnitude of Bus PTIS Questionnaire expansion factors.

Table 13 – Representative Populations for Train Expansion Stages

Expansion Factor	Expansion Step	Outbound		Inbound (Reversed)	
		Mean Expansion Factor	Population	Mean Expansion Factor	Population
-	Questionnaires	-	4,635	-	2,151
EF01	Adults	1.67	7,725	1.62	3,481
EF02	Capping EF01	-	-	-	-
EF03	Accompanied Children	1.04	8,017	1.03	3,601
EF04	Unaccompanied Children	1.12	9,000	1.11	3,998
EF05	Time Bias Factor	-	-	-	-
EF10	Expansion by Time Period	-	-	-	-
EF06	Missed Pax (in corridor & time period)	-	-	-	-
EF07	Off-peak passengers	1.06	9,523	1.06	4,231
EF08	Missed Pax (in region)	-	-	-	-
EF09	Missed Pax (unknown region)	-	-	-	-
EF11	Trip reversal	-	-	2.25	9,523
Overall		2.05	9,523	4.43	9,523
Grand Total		2.81	19,045		