
Appendix C: Methodology

Introduction

This appendix describes various methodological aspects of the Longitudinal Immigration Survey: New Zealand (LisNZ). The target population, sampling frame, and survey populations are described in the first section. The second section describes the sampling design, including the allocation of the sample across the various sub-populations of interest. Non-response is discussed in the third section. The weighting process used to obtain estimates from the survey is described in the fourth section. The last section discusses the quality of the data.

Survey population and sampling frame

The target population for the LisNZ consisted of all migrants (excluding refugees) who:

- were approved for residence in New Zealand from 1 November 2004 to 31 October 2005
- were aged 16 years or over at the time of residence approval
- were already in New Zealand at the time of residence approval or arrived in New Zealand within 12 months of residence approval
- spoke one of the seven designated survey languages.

Consequently, the target population excluded temporary visitors and all people from Australia, Niue, the Cook Islands, and Tokelau. Migrants from Australia were excluded because they are entitled to enter New Zealand without applying for a residence permit or visa. In addition, migrants from Niue, the Cook Islands, and Tokelau were excluded, because people from these countries have automatic rights to New Zealand citizenship. Refugees were excluded from the target population because their routes to permanent residence, as well as their settlement experiences, are very different from those of other migrants.

For the period covered by the LisNZ (1 November 2004 to 31 October 2005), over 40,000 migrants were granted permanent residence in New Zealand.

The survey population, from which the sample was selected, consisted of migrants in the target population who were living in the North Island, South Island, or Waiheke Island at the wave 1 interview date,¹ and who could understand at least one of the designated survey languages (English, Mandarin, Cantonese, Samoan, Korean, Hindi, or Punjabi). The survey population also excluded migrants who had left New Zealand permanently at the wave 1 interview date.

The sampling frame for the LisNZ was constructed from the Immigration New Zealand administrative database known as the Application Management System. This database is a 24-hour global processing system for permanent residence applications that contains the following information about the applicants:

- personal information, including name, address, date of birth, sex, and nationality
- details of the permanent residence application such as date of approval, residence approval category, type of visa or permit issued, and whether the applicant is a principal or secondary applicant.

¹ Wave 1 interviews were conducted at six months after arrival for offshore migrants and six months after residence approval for onshore migrants.

Information was extracted daily from the database to construct the sampling frame. The extraction process involved:

- checking stratification variables to ensure these variables were not missing
- choosing a unique ‘identity’ for each migrant, especially when a migrant’s details were updated or corrected in the database
- dealing with migrants approved in different applications on the same day
- updating information on the sampling frame.

In addition to the daily extract, a weekly extract included all migrants approved for residence during the week. This process may have resulted in migrants changing strata or changes to the number of migrants within an application. In such cases, the information about the migrants was updated in the sampling frame but their probabilities of selection were always based on the original information, unless additional migrants were identified in an application.

Sampling design

Longitudinal survey

The survey has a longitudinal design where migrants are interviewed three times (waves 1, 2, and 3) over three years: at around six months, 18 months, and 36 months after residence approval for onshore migrants, or six months, 18 months, and 36 months after arrival for migrants approved offshore.

The migrants who responded to the wave 1 interview will be followed up for interview in waves 2 and 3. Migrants who do not complete an interview in any wave will not be followed up in subsequent waves.

The objective of the survey was to produce reliable estimates for a number of sub-populations defined by immigration approval category and region of origin. A stratified systematic sampling design was used to achieve this. The various aspects of the design are described below.

Stratification

There were two stages of stratification. The first stage of stratification divided the survey population into 10 superstrata, which were defined by the residence approval categories. Many of the superstrata were stratified further by location of approval (onshore or offshore), region of origin, and type of applicant (principal or secondary), making a total of 40 design strata.

Sample allocation

The survey was designed to produce estimates for several sub-populations defined by immigration approval category and region of origin. It was estimated that 5,000 completed interviews in wave 3 would produce estimates of the required accuracy, with sample sizes of between 100 and 500 respondents in each design stratum.

The requirement to produce regional estimates within the Skilled Migrant Category and Family Partner category meant that the smaller regions were over-sampled within these categories. Proportional allocation was applied at the level of application type (offshore or onshore). When it was not possible to achieve proportional allocation, then all the offshore migrants were selected and the remainder of the sample was selected from onshore migrants. The final target allocation of the sample across strata is shown in Table C1.

Table C1: Target sample allocation

Residence approval category	Region of origin							Total
	ESANA	UK/Irish Republic	North Asia	South Asia	South East Asia	Pacific	Other	
Business – Principal								300
Business – Secondary								150
General Skills								300
Talented Migrants ⁽¹⁾								150
Skilled Migrants	350	500	350	200	150	350	300	2,200
Family Partner	300 ⁽²⁾		150	50	150	150	150	950
Family Parent								300
Family Quota								300
Samoan Quota								150
Pacific Access								100
Other								100
Total								5,000
Allocation to visa								
Business – Principal								165
Business – Secondary								76
General Skills								0
Talented Migrants ⁽¹⁾								0
Skilled Migrants	44	180	70	14	13	61	56	438
Family Partner	131 ⁽²⁾		34	13	80	46	23	327
Family Parent								231
Family Quota								207
Samoan Quota								150
Pacific Access								100
Other								21
Total								1,715
Allocation to permit								
Business – Principal								135
Business – Secondary								74
General Skills								300
Talented Migrants ⁽¹⁾								150
Skilled Migrants	306	320	280	186	137	289	244	1,762
Family Partner	169 ⁽²⁾		116	37	70	104	127	623
Family Parent								69
Family Quota								93
Samoan Quota								0
Pacific Access								0
Other								79
Total								3,285

Notes: ⁽¹⁾ Includes the Talent and the Long Term Skills Shortage List Occupation policies.

⁽²⁾ Includes both ESANA (Europe, South Africa, and North America) and UK/Irish Republic.

Sample selection

A systematic sample of migrants was selected within strata according to the allocations in Table C2. A large initial sample was required to achieve 5,000 completed interviews at wave 3, after allowing for non-contact, non-response, and attrition.

Sampling fractions within the 40 strata were based on the estimated number of migrants that would be approved over the 12-month sampling period. In order to control the final achieved sample sizes, sampling fractions within some strata had to be varied over the sampling period.

A systematic sample selection method was used mainly to control the number of migrants selected from the same application. In cases where more than two migrants were selected from the same application, only two migrants were retained in the sample, with different rules applied to applications in the Business category and the rest of the residence approval categories. All principal applicants within the Business category were selected, and if more than two migrants were selected from the same application, the principal applicant was retained together with one randomly chosen secondary applicant.

If more than two migrants were selected from any other approval categories, two migrants were randomly chosen to remain in the sample. The second step was applied across strata, as it was possible to select migrants from the same application in different strata. For example, migrants in the same application may have different nationalities, so they would be allocated to different strata in the survey population.

Table C2: Sample selected at wave 1

Residence approval category	Region of origin							Total
	ESANA	UK/Irish Republic	North Asia	South Asia	South East Asia	Pacific	Other	
Business – Principal								673
Business – Secondary								335
General Skills								320
Talented Migrants ⁽¹⁾								313
Skilled Migrants	852	1,388	969	775	681	339	411	5,415
Family Partner	742 ⁽²⁾		336	305	321	346	94	2,144
Family Parent								887
Family Other								833
Samoan Quota								553
Pacific Access								496
Other								233
Total								12,202

Notes: ⁽¹⁾ Includes the Talent and the Long Term Skills Shortage List Occupation policies.

⁽²⁾ Includes both ESANA (Europe, South Africa, and North America) and UK/Irish Republic.

Non-response

Despite all the efforts in tracking the migrants selected for wave 1 interviews, some migrants were not able to be interviewed or provide answers to all the questions during the interview. This section summarises the different levels of non-response, namely unit and item non-response.

Unit non-response

Unit non-response applies to migrants who were selected in the wave 1 sample but were not interviewed. Their contributions are included in the adjustments made to the respondents' sampling weights. This unit non-response adjustment is discussed in the section 'Weight adjustment for calibration'. It is based on a migrant's eligibility status as described in Table C3.

Table C3: Response status for wave 1 sample

Eligibility status	Number in the sample			Percent (%)		
	Offshore	Onshore	Total	Offshore	Onshore	Total
<i>Ineligible</i>						
Did not arrive in time or take up residence	145	0	145	14.4	0.0	10.8
No initial contact in New Zealand	755	229	984	74.8	68.0	73.1
Other reason ⁽¹⁾	109	108	217	10.8	32.0	16.1
Total	1,009	337	1,346	100.0	100.0	100.0
<i>Eligible</i>						
Response	1,937	5,200	7,137	57.0	69.7	65.7
Cannot be contacted	1,336	1,796	3,132	39.3	24.1	28.9
Refusal	95	405	500	2.8	5.4	4.6
Bilingual interviewer not available or respondent was sick	32	55	87	0.9	0.7	0.8
Total	3,400	7,456	10,856	100.0	100.0	100.0
Total number in the sample	4,409	7,793	12,202			

Notes: ¹ Includes migrants who died before the interview date, migrants who did not speak a survey language, and migrants aged under 16 at the time of residence approval.

From the sample of 12,202 migrants selected in wave 1, 217 were not eligible to take part in the survey, 145 did not arrive in New Zealand in time, and 984 had no initial contact address in New Zealand. Of the remaining 10,856 migrants, 7,137 were interviewed. This corresponds to a 66 percent response rate.

Most of the non-response was due to non-contact (84 percent) rather than respondent refusal. Non-contact was particularly high for migrants who were approved offshore, where the response rate was 57 percent compared with 70 percent for those approved onshore.

Table C4 shows the number of respondents by region and approval category. The numbers have been rounded to protect confidentiality, with cell counts of 1 to 10 being rounded up to 10 and all other cell counts rounded to the nearest 5.

Table C4: Wave 1 response data

Region of origin	Immigration approval category							Total
	Skilled principal	Skilled secondary	Business category	Family Partner	Family Parent	Pacific category	Other	
UK/Irish Republic	590	420	50	290	90	...	60	1,500
South Africa	170	150	10	15	40	...	25	415
North America	85	45	10	85	10	10	10	235
Rest of Europe	105	80	10	85	10	...	20	310
North Asia	455	195	285	195	115	...	90	1,335
South Asia	395	195	15	195	75	...	25	905
South East Asia	320	185	10	180	15	...	40	750
Pacific	145	105	30	195	90	465	180	1,215
Other	195	145	10	60	10	...	45	470
Total	2,460	1,525	430	1,310	450	465	495	7,135

Notes: ... Not applicable.

Item non-response

Item non-response corresponds to incomplete information from the migrants who were interviewed in wave 1. Imputation methods are usually used to fill in these missing items. However, imputation was not applied in LisNZ wave 1 as the item non-response rate was less than 1 percent for most variables in the survey.

Estimation

This section describes the sampling weights and other adjustments that have been applied to obtain estimates from the LisNZ.

Design weight

A design weight was attached to each migrant in the LisNZ sample to reflect the probability of being selected in the sample. The design weights were different across strata because of unequal probabilities of selection. Furthermore, the design weights within strata varied over the 12-month sampling period because the sampling fractions were changed throughout the sampling period to control the final achieved sample size.

An initial 'application' adjustment was made to the design weight to account for the retention of a maximum of two migrants per application. The application adjustment factor had a value of one if the application had at most two migrants in the sample. If an application from the Business Category had more than two migrants in the sample, then the application adjustment factor for the principal applicant was equal to one, while the application adjustment factor for the secondary applicant equalled the number of migrants in the application minus one. Finally, for an application from any other residence approval categories that had more than two migrants, the application adjustment factor was equal to the number of migrants in that application divided by two.

A second adjustment (a 'stratum' adjustment) was made to the 'application' adjusted weights to ensure the sum of the stratum-adjusted weights within a stratum was the same as the total number of migrants in the survey population from that stratum.

Weight adjustment for unit non-response

A 'unit non-response' adjustment was applied to the stratum-adjusted weight to account for the unit non-respondents based on weighting classes. These weighting classes were defined with the intention that the characteristics of the respondents and non-respondents would be similar. For wave 1, the weighting classes were based on the 40 strata. In four of the strata (Talent, General Skills, Samoan Quota, and Pacific Access Category), the weighting classes were defined by type of application (offshore or onshore). Hence, 44 weighting classes were used for 'unit non-response' adjustments in wave 1.

Weight adjustment for calibration

A final adjustment was applied to the 'unit non-response' adjusted weights to benchmark to known population totals. The population totals used for benchmarking were the actual number of migrants approved over the survey reference period by strata, sex, and age groups. For wave 1, a population of 37,633 migrants was used as the benchmark, of which 36,223 were deemed eligible for the survey.

Rounding and suppression of estimates

The weighted estimates provided in this report have been randomly rounded to base 10 and the percentages have been calculated using the rounded values. As a result of rounding, the estimated totals may differ from the sum of the individual cells in a given table.

Cells with weighted estimates of fewer than 20 people have been suppressed for confidentiality protection of the respondents. These cells, as well as the proportions based on them, appear as 'S' (suppressed) in the tables.

Data quality

Two types of error are possible in estimates based on a sample survey: sampling error and non-sampling error. The results from the LisNZ are subject to both of these sources of errors, and this should be taken into consideration when analysing the results from the survey.

Sampling errors

Sampling error is a measure of the variability that occurs because information has been calculated from a sample of migrants rather than the entire population of migrants in a given reference period.

The LisNZ used a stratified sampling design and the sampling errors are somewhat greater than those that would have been obtained if a simple random sampling design had been used. Three aspects contributed to this. First, the requirement to produce regional estimates within the Skilled Migrant Category and Family Partner Category meant smaller regions were over-sampled. Secondly, sampling fractions within some strata varied over the sampling period, and thirdly non-response adjustments. Interviewing only two migrants from the same application may also have some effect on increasing the sampling errors.

Sampling errors have been estimated using a replication variance estimation method, in particular the delete-a-group jackknife variance estimation method.²

² P S Kott. 'Using the delete-a-group variance estimator in practice.' In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 1998, pp 763-768.

The sampling errors for a selection of variables from the LisNZ are in Appendix D. Design effects have been calculated which quantify the increase in sampling error due to the sampling design relative to a simple random sampling design. In the LisNZ, the population groups, defined on the basis of immigration approval category and region of origin, have average design effects ranging from 0.98 to 1.29 (see Table C5), based on the selection of variables in Appendix D.

Table C5: Average design effects

Immigration approval category	Average design effect	Region of origin	Average design effect
Skilled principal	1.29	UK/Irish Republic	1.09
Skilled secondary	1.19	South Africa	1.17
Business category	1.10	North America	1.00
Family Partner	0.98	Rest of Europe	1.14
Family Parent	1.14	North Asia	1.22
Pacific	1.11	South Asia	1.09
Other	1.00	South East Asia	1.17
Total	1.19	Pacific	1.09
		Other	1.12
		Total	1.23

The average design effects for a population group in Table C5 can be used, in conjunction with the sample size and sampling errors based on a simple random sample, to obtain indicative estimates of sampling error for other estimates contained in this report.

Non-sampling errors

Non-sampling errors include non-response bias, inaccuracies in the responses of the migrants during the interviews, and errors made during the processing of the data. Statistics New Zealand applies survey monitoring procedures, such as editing of the data collected from the interviews, to minimise these types of errors, but they may still occur and are difficult to measure.