Defining migrants using travel histories and the '12/16-month rule'

Contents

Tables and figures

List of tables

List of figures

Introduction to the new measure

Purpose

This report describes a new measure of determining the contribution of international migration to changes in New Zealand's resident population. The measure uses linked travel histories and a rule for determining a change in resident status. This rule is independent of the individual's legal residence status and also independent of the information stated on arrival and departure passenger cards.

We provide summaries of migrant arrivals and departures classified by the '12/16 month rule', and include some comparisons with the official released figures of permanent and long-term (PLT) arrivals and departures.

We describe the operational rules underlying the 12/16-month rule for classifying migrant status. The paper also outlines the data integration process of border movements that is necessary for creating individuals' travel histories, including a quality assessment of this process.

Background to the new measure

Observing individual passengers' sequences of border movements over time presents an opportunity for measuring changes in resident status that are based on actual patterns of stay in, or absence from, New Zealand.

Applying a measurement rule to individuals' travel histories over time, to classify their migrant status, may produce different results from the passenger-type class determined by information they initially record on the passenger cards.

Based on passenger cards, currently, a person may transition from being reported as an 'overseas visitor arrival' to being reported as a 'New Zealand-resident departure traveller' if their stay is eventually 12 months or longer. For example, when the person transitions onshore from a short-term permit on arrival to extended permits or subsequent longer-term temporary resident-visa approvals. By doing this, they satisfy the period of residence in New Zealand required for classification as a New Zealand resident. Similarly, a New Zealand resident intending to make a short-term overseas trip may decide to live abroad for a longer time period than they report on their departure card.

Differences between actual travel histories, and reported intended stay or time away from New Zealand on the passenger cards that result in shorter or longer stays or absences, may affect the relevance or accuracy of the initial classification of migrant movements (using the passenger card information). This can result in estimates of migrant arrivals and departures being over- or under-reported when the passenger card is the primary information source used to determine migrant status.

Integrating passengers' border movements over time enables us to observe their actual patterns of stay in New Zealand. An appropriate measurement rule applied to their travel histories allows us to classify migrant status as well as determine other passenger type categories (eg overseas visitor or New Zealand-resident traveller).

The new measure also introduces added opportunities for us to extend the measures of international migration statistics. For example, there may be interest in alternative measures of short-term as well as long-term migration. Further, traveller histories represent a longitudinal data source that enables other statistical measures of migration (eg return migration and step-migration).

Integration of individuals' histories of migrant arrivals and departures classified by the 12/16-month rule with Stats NZ's Integrated Data Infrastructure (IDI) will also facilitate other analysis of migration trajectories and outcomes.

Summary of new data series and measure

We prepared a historical series of migrant arrivals and departures classified by the 12/16-month rule for July 2001 onwards. The classification method uses a month as a reference period. Any traveller with at least one border movement in a given month has their resident status reviewed. The combined information on the assumed resident status at the start of the month, the direction of the last movement in the month, and the 16-month follow-up travel history, determines the traveller's final migrant status.

We used regular releases of border movements data, including associated movement and passenger identities, to Stats NZ for production of a historical series of estimated migrant arrivals and departures using the 12/16-month rule. Individuals' travel histories were prepared for all border movements from 1999 onwards. The new classification method requires a 'start-up' period to allow a build-up of archival resident status changes.

A data quality assessment of the integrated border movements indicated results that were comparable with results from other main data integration projects undertaken by Stats NZ, such as the IDI.

Migrant arrivals and departures by the 12/16-month rule: examples

In this section we compare historical estimates of migrant arrivals and departures classified by the 12/16-month rule with the official series of PLT arrivals and departures (Statistics NZ, Migration statistics) as measured by the passenger cards.

Migrant arrivals and departures classified by the 12/16-month rule are consistently estimated at higher levels than released figures based on the passenger card information (figure 1). Passengers' reported intentions of staying or leaving permanently are under-estimating the actual outcomes. From 2012 onwards, there was a consistently smaller difference between estimated migrant flows using the 12/16-month rule and the official series when compared with earlier years.

Figure 1

Migrant arrivals and departures
By 12/16-month rule and PLT measures
December 2001–14 years
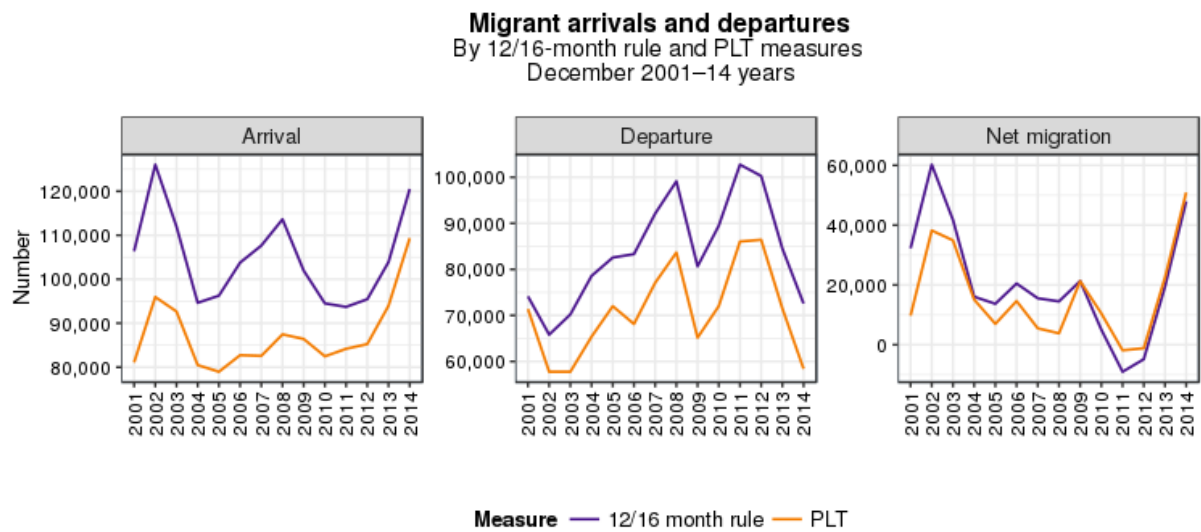
Source: Stats NZ

Comparing net migration estimates for the two definitions shows the 2001–08 period was characterised by an under-estimation of net migration gains by the PLT series (figure 2); it peaked at 22,500 in 2001. For 2010–14, net migration gains by the PLT series were over-estimated; it peaked at 7,200 in 2011 but this gap was consistently estimated at around 3,000 a year in 2013–14.

Figure 2

Difference in net migration between 12/16-month rule and PLT measures
December 2001–14 years
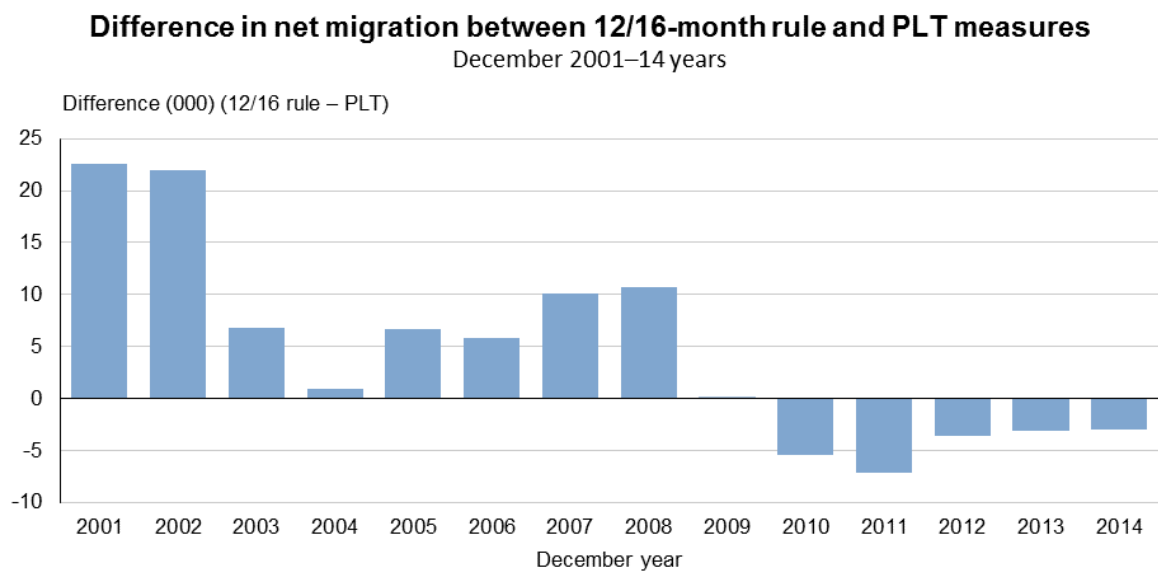
Source: Stats NZ

Patterns of under-estimation of total migrant arrivals and departures based on the passenger card information differ by country of citizenship, and by border entry visa. Passengers who are New Zealand citizens arriving in New Zealand are often likely to report short-term stay intentions, while the actual outcomes, shown by a 16-month follow-up period, suggest they have returned to live in New Zealand (figure 3). For this reason, estimates of migrant arrivals of New Zealand citizens by the 12/16-month rule are consistently higher than the official series; this averaged 4,000 a year in 2009–14. In contrast, PLT departures of New Zealand citizens are much closer to the outcomes estimated by the 12/16-month rule.

Figure 3

Migrant arrivals and departures by selected country/region of citizenship
By 12/16-month rule and PLT measures
December 2001–14 years
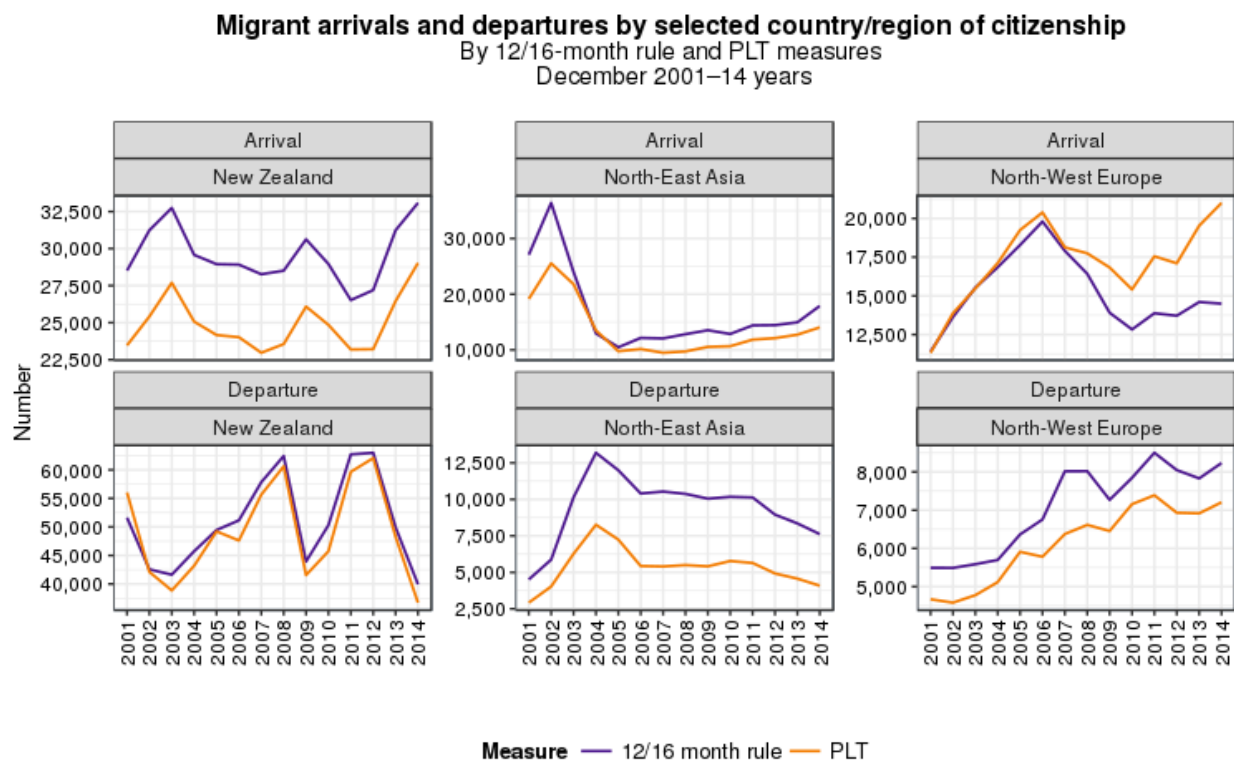
Source: Stats NZ

North-East Asian (eg China, Korea, Taiwan, and Japan) migrant flows show the opposite pattern to New Zealand citizens. Their inflows measured by the 12/16-month rule are generally close to the intentions measured by the PLT series, but with the PLT series giving a consistent under-estimation of migrant outflows. For North-West European citizens (including UK and Ireland), comparing migrant flows by the two measures show an increasing higher estimate of migrant arrivals from 2010 onwards by the PLT measure. In contrast, migrant departures of these citizens were likely also to be under-estimated by the official series.

Figure 4

Migrant arrivals by selected border entry visa type
By 12/16-month rule and PLT measures
December 2001–14 years
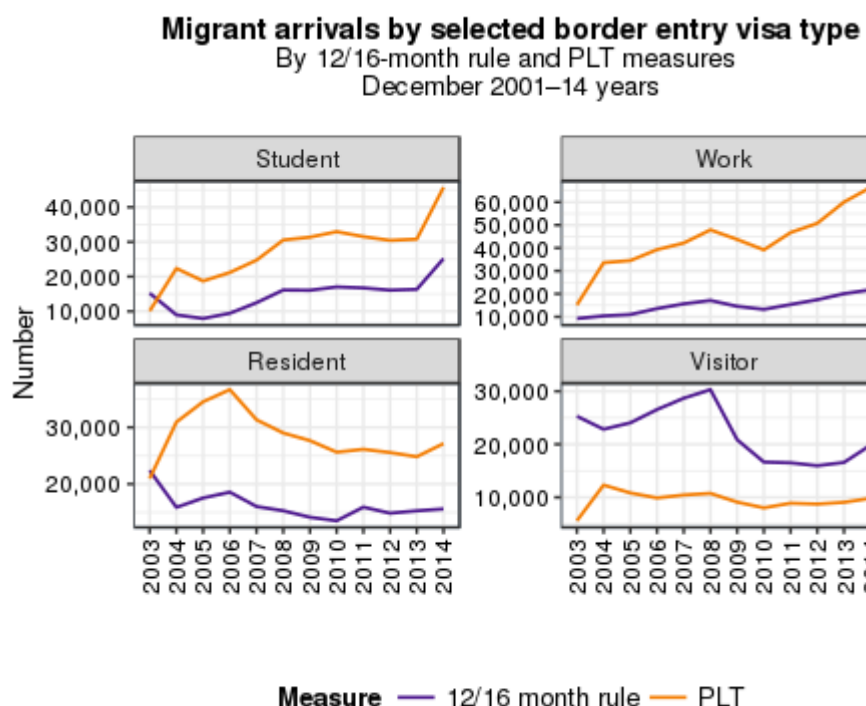
Source: Stats NZ

Passengers arriving with an approved work visa or student visa may report their intention to stay in New Zealand permanently on the passenger card and be assigned PLT migrant status at the time of arrival in New Zealand. However, many choose to leave New Zealand early (eg within a year) and by the 12/16-month rule they do not gain migrant status (figure 4). On the other hand, passengers holding a visitor visa are likely to indicate a short-term stay on arriving in New Zealand, but many will renew their visa or transition to other types of visa on-shore.

Alternative estimates of net migration

Estimates of migrant arrivals and departures using the 12/16-month rule provide an alternative measure of changes in the resident population due to migration. These can be compared with previously published estimates:

- **PLT** using passenger card information (Statistics NZ, nd)
- **total passenger movements** using passenger card information – while the large volume and seasonality of arrivals and departures makes this an inappropriate migration measure in the short term, it is more useful over periods of several years (Statistics NZ, nd)
- **estimated net migration using census-based population estimates** – implied net migration, and independent of passenger card information, is the residual after subtracting births and adding deaths (Statistics NZ, nd).

Migration estimates measured by the 12/16-month rule generally align with these alternative estimates (figure 5). However, the 12/16-month estimates provide further evidence that the PLT measure under-estimated net migration during the early 2000s. In the more-recent intercensal period (2006–13), the alternative net migration estimates were much closer to each other.

Figure 5

Alternative estimates of net migration
Year ended June, 2002–06 and 2007–13

Source: Stats NZ
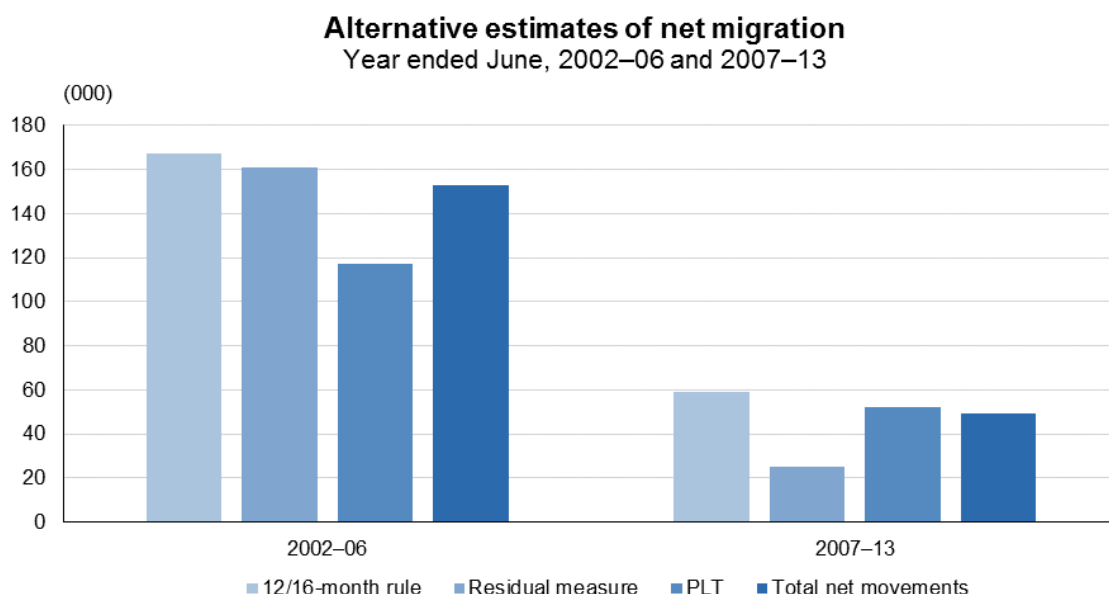
Change to definition of migrant status

Implementing a measurement rule for estimating migrant status based on actual stay in, or absence from, New Zealand is likely to reduce the magnitude of a revision required in intercensal population estimates series. When a new base estimated resident population (ERP) is released, following the most-recent census, we revise population estimates to account for the discrepancy observed in the census year estimates.

Historical releases of updates to the ERP during the intercensal period have not included adjustments for the net migration component. A major component of the discrepancy is likely to be the net effect of deficiencies in estimating changes in resident status at the time of international travel (Statistics NZ, 2016).

See How accurate are population estimates and projections?

Under the 12/16-month rule, passengers who are eligible to be counted as part of the ERP are unchanged. However, applying the new measurement rule using passengers' travel histories implies a different definition of when a person has a final status of being included in or excluded from the resident population.

Definition of migrant status

Below are the differences between the resident status of the traveller and the classification process to assign migrant status – given by the PLT classification and by the 12/16-month rule, respectively.

Using passenger card and retrospective travel history:

- New Zealand resident status at point of travel – report 'living' or 'not living' in New Zealand for 12 months or more, and confirmed by travel history for PLTs

- Classification of migrant status – report intend to stay in, or be absent from, New Zealand for 12 months or more.

Using 12/16 month rule applied to linked travel history:

- New Zealand resident status at point of travel – determined from passenger classification history in previous 16 months

- Classification of migrant status – total duration of stay(s) in, or away from, New Zealand is 12 months or more during 16-month follow-up period.

The 12/16-month rule

Classifying migrant status by the 12/16-month rule requires us to classify passengers' past-border movements, and their travel sequences observed from movements over a 16-month follow-up period. Subsequently, after a 16-month travel history has been established, the 12/16-month rule assigns a final migrant status for travellers with an observed change in resident status. (Appendix 1 summarises the classification process flow.)

Assumptions we make about resident status for travellers at the start of their first movement in a given reference month use classification histories from the previous 16 months of processed passenger movements (table 1(a)). If travellers have not had border movements during that period, the direction of their first movement defines their resident status at the start of this period.

Table 1

The 12/16-month rule

1 (a): Estimated resident status before first movement in reference month

Latest resident update in 16-month rolled forward archive	Direction of first movement	Assigned resident status
Resident	…	Resident
Non-resident	…	Non-resident
None	Arrival	Non-resident
None	Departure	Resident

1 (b): Assigning resident status following movement in reference month

Total duration of stay(s) in NZ (days) over 16-month follow-up period	Assigned resident status
<= 121	Non-resident
>121 to < 365	Same as resident status assigned at start of month (1 (a))
>= 365	Resident
Note: 4 months are taken as 121 days, and 12 months as 365 days. Symbol: … not applicable Source: Stats NZ	

The resident status we assign to the traveller ID before the first movement is re-evaluated following the observed travel sequence in the reference month. From observed travel sequences over the 16-month follow-up period, we calculate total duration of stays in New Zealand from each point of travel in the reference month to the end of the follow-up period.

Resident status is assigned based on the total duration of stay(s) in New Zealand (table 1(b)). For example, if the total duration sums to 122–364 days, the resident status we assign to this movement is the same as the resident status assigned to the traveller before the first movement in the reference month.

Finally, the classification process determines the migrant status for travellers with a changed resident status by incorporating the direction of the last movement in the reference month (table 2). We assume travellers classed as migrant arrivals are non-resident at the start of the month. Following their last movement, an arrival, they are assigned resident status. Equivalently, if travellers' last movement is a departure with non-resident status, and they are assigned resident status at the start of the month, they are classed as migrant departures. New Zealand-resident travellers and overseas visitors are short-term passengers who do not contribute to changes in New Zealand's resident population.

Table 2

Classification of passenger type by the 12/16-month rule

For passengers with a change in resident status during the reference month

Resident status at start	Resident status at end	Direction of last movement	Passenger type
Non-resident	Resident	Arrival	Migrant
Resident	Non-resident	Departure	Migrant
Resident	Non-resident	Arrival	NZ-resident traveller
Non-resident	Resident	Departure	Overseas visitor
Source: Stats NZ			

For the purpose of classifying a historical series of border movements using the 12/16-month rule, the first 16 months of processing represent a data initialisation phase. This phase aims to build-up a 16-month rolled-forward archive of passengers' classification histories and their assigned resident status. The data initialisation phase assigns resident status by observing a 16-month travel sequence both before and after the reference month.

The data: travel histories

This section describes the data integration process we used to create individuals' travel histories. It also assesses the accuracy of the historical series of estimated migrant arrivals and departures released with this report (see data tables in 'Available files').

[Defining migrants using travel histories and the 12/16-month rule data collection methodology](#) has the metadata for these data tables.

Border movement data and all movement identities were accessed at the unit record level. For the specific data integration process, outlined below, we used the added content of names and complete date of birth information. All border movements from November 1998 onwards entered the data integration process. Identifying aircraft crew movements in the historical border movements data is not possible. However, we assume they have minimum impact on the final estimates of migrant arrivals and departures using the 12/16-month rule.

Data integration of border movements

A probabilistic data integration process of border movements serves to assign a unique ID to all international travellers. The process applies the Fellegi-Sunter methodology (Statistics NZ, 2015) over a three-pass linking process. The first pass is a deterministic linking step using the movement record identifier derived from passport identifiers at the time of travel. This is followed by two probabilistic linking steps: one takes care of travellers who change passport nationality or use more than one passport, and the second accounts for travellers changing their surname. The blocking and linking variables and other technical descriptions of the record linking methodology are in appendix 2.

Data quality assessment

The data quality of the travel histories is partly measured by the link rate of the data integration process and partly by assessing link accuracy. The entire data integration of border movements from November 1998 onwards created 28 million unique traveller IDs and the overall achieved link rate was 98 percent.

We measured link accuracy by estimating the proportion of incorrect links occurring in the data integration process (false positive rate). Using a sampling approach, we estimated the false positive rate for each linking pass. The overall false positive rate and the sampling error were comparable to results in other Stats NZ data integration projects (see appendix 3).

Imputation for missing movements

We included an imputation step to ensure unique and plausible travel histories were input to the classification process by the 12/16-month rule. A very low and decreasing imputation rate over time indicates a negligible effect on the accuracy of estimated migrant movements.

The data integration processes to create the movement record identifier in the IDI, and to create the unique traveller ID, both resulted in some individuals' having implausible travel sequences. This means that a few individuals' travel sequences may have missing arrivals and/or departures. The imputation step estimated the missing border movement date-times as the mean date-times of the two sequential arrivals or departures.

The proportion of missing movements decreased over time. This resulted from ongoing quality improvements in the electronic capture of border movements. Results from the imputation step indicated that on average, from 2010 onwards, the rate of imputed movements was less than 0.15 percent for both annual arrivals and departures. Earlier annual imputation rates were between 0.15 and 0.22 percent.

Classification of PLT migrants

Current processes for classifying passenger type relies on travellers' self-reported information on the passenger card at the time of travel; i.e. their intended length of stay or absence and whether they have lived in New Zealand in the last 12 months. The process also uses linked passenger movements for any final editing of the passenger type for people stating they are permanent long-term (PLT) arrivals or departures.

We classify a traveller as a PLT arrival if, at the time of arrival, they state they intend to live in New Zealand for 12 months or more and their travel history confirms they were not living in New

Zealand in the 12 months before travel. From the travel history, we assess the person's actual length of stay in New Zealand before the time of travel. When the traveller has spent less than 12 months in New Zealand during the last 16 months we assume they were not living in the country before the time of arrival.

Similarly, we classify a traveller as a PLT departure if, at the time of departure they state they intend to be absent for 12 months or more and their travel history confirms they lived in New Zealand in the past 12 months.

Travel history informs us whether the person has spent at least 12 of the last 16 months in New Zealand before the time of departure, and is therefore a New Zealand resident.

Conclusion

Estimation of migrant arrivals and departures using the 12/16-month rule provides a different measure of changes in the resident population that are due to migration. This estimation does not rely on the passenger card information at all. Since the method requires us to observe follow-up travel histories, there is a 16-month lag following a reference month before the final estimates by this rule become available.

Evidently, estimates of migrant arrivals and departures by the two measures, PLT and 12/16-month rule, are highly correlated (figure 1). For this reason, the PLT series, based on the passenger card information, represent a timely estimate of migrant arrivals and departures. The net PLT estimate provides an early indication of the contribution of migration to changes in the resident population.

Releases of final estimates of migrant arrivals and departures by the 12/16-month rule rely on timely updates to the archived traveller IDs. Updates to traveller IDs, and their associated classification entities, will enable integration with the IDI travel and migration datasets (Stats NZ, 2017). The ability to combine individuals' travel histories with other IDI information, such as immigration approval decisions and socio-demographic variables, for analysis of settlement characteristics, will be useful to inform future migration policy changes.

Implementing the 12/16-month rule for estimating changes in New Zealand's resident population from net migration will also lead to a harmonised statistical definition of migrant status with Australia. The Australian Bureau of Statistics (ABS) has, from mid-2007, released quarterly series of final estimates of net gains to the Australian resident population from overseas migration using the 12/16-month rule (ABS, 2007). This introduces opportunities for bilateral analysis of trans-Tasman migration flows.

References

Australian Bureau of Statistics (2007). Statistical implications of improved methods for estimating net overseas migration. 3107.0.55.005 - Information paper. Retrieved from abs.gov.au.

Statistics NZ (nd). International travel and migration data collection methodology. Retrieved from www.stats.govt.nz.

Statistics NZ (nd). Migration statistics. Retrieved from www.stats.govt.nz.

Statistics NZ (2015). Data integration manual.

Statistics NZ (2016). Integrated data infrastructure. Retrieved from www.stats.govt.nz.
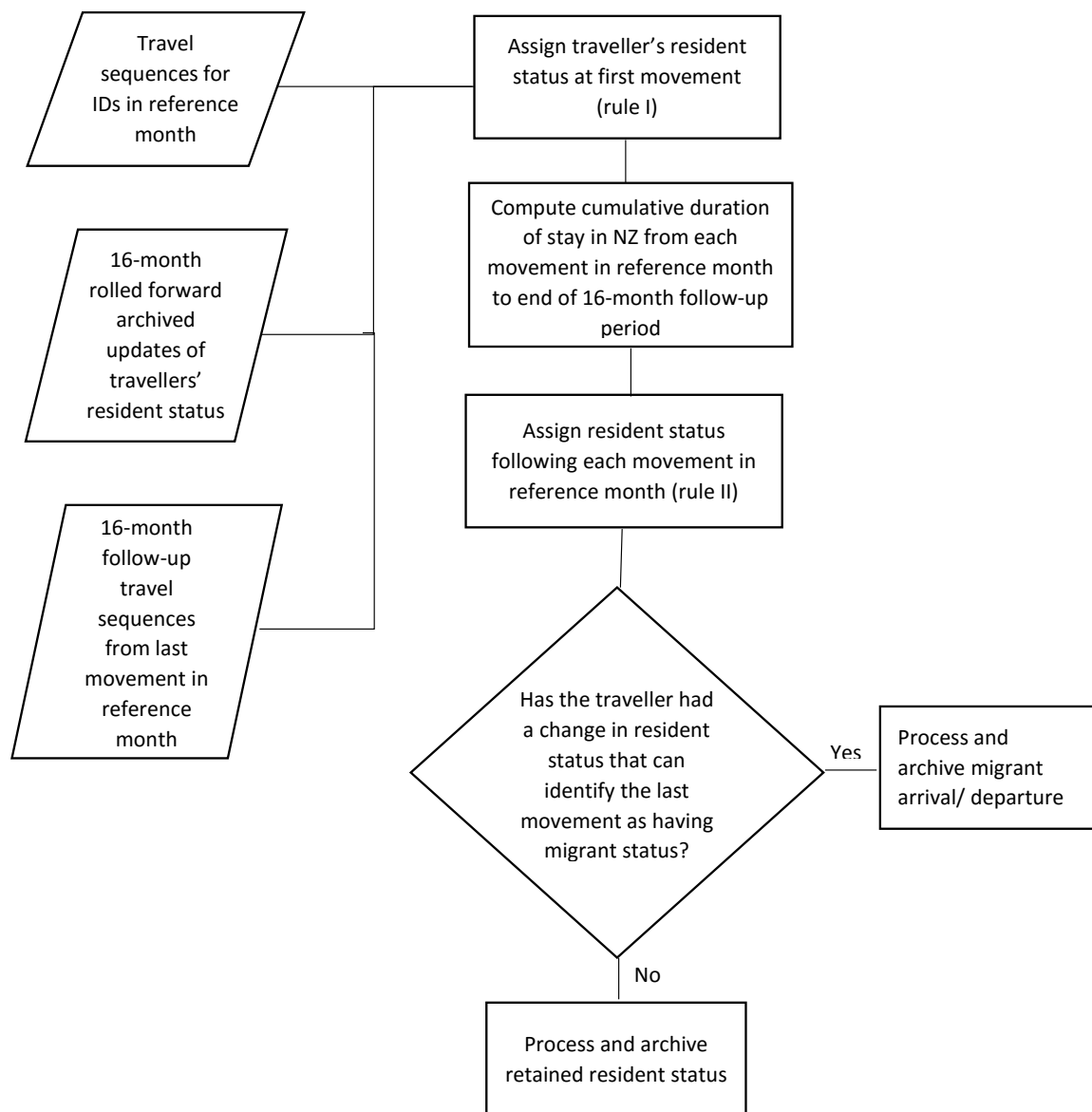
Statistics NZ (2016). How accurate are population estimates and projections? Retrieved from www.stats.govt.nz.

Stats NZ (2017). Integrated data infrastructure: Travel and migration data. Retrieved from www.stats.govt.nz.

Appendix 1: Process flow for the 12/16-month rule

Figure 6

Classification process flow for travellers with border movements in a given reference month

```
┌─────────────────┐                    ┌──────────────────────┐
│ Travel          │                    │ Assign traveller's   │
│ sequences for   │                    │ resident status at   │
│ IDs in reference│                    │ first movement       │
│ month           │                    │ (rule I)             │
└─────────────────┘                    └──────────────────────┘
                                                  │
┌─────────────────┐                    ┌──────────────────────┐
│ 16-month        │                    │ Compute cumulative   │
│ rolled forward  │                    │ duration of stay in  │
│ archived        │                    │ NZ from each movement│
│ updates of      │                    │ in reference month   │
│ travellers'     │                    │ to end of 16-month   │
│ resident status │                    │ follow-up period     │
└─────────────────┘                    └──────────────────────┘
                                                  │
┌─────────────────┐                    ┌──────────────────────┐
│ 16-month        │                    │ Assign resident      │
│ follow-up       │                    │ status following each│
│ travel          │                    │ movement in reference│
│ sequences       │                    │ month (rule II)      │
│ from last       │                    └──────────────────────┘
│ movement in     │
│ reference       │
│ month           │
└─────────────────┘
```

Has the traveller had a change in resident status that can identify the last movement as having migrant status?

Yes → Process and archive migrant arrival/ departure

No → Process and archive retained resident status

Source: Stats NZ

Appendix 2: Data integration methodology

This appendix describes the record linking process of all border movements from November 1998 onwards. The process assigns a unique traveller ID that is associated with the border movements. This allows us to observe individuals' travel histories and compute actual length of stays in New Zealand as input to classifying migrant status using the 12/16-month rule.

Data

Two datasets received from New Zealand Customs Service are the movements (ie each movement in or out of New Zealand) and the movement identities datasets (identity information of the person associated with the movement – from passport information). The movement dataset includes additional information, such as the direction and date of the movement as well as the embarkation and disembarkation ports.

Linking methodology

The linking process follows the Fellegi-Sunter methodology and uses a de-duplicate independent match in the IBM QualityStage data integration software. It looks for duplicates in the data by identifying groups of records that share common attributes, and allows the user to correct, merge, or eliminate the duplicate entries. When comparing two records it assigns a field weight – a measure that reflects how similar they are.

Linking and blocking variables

The end product of a data integration process is to create links of records. Two records are a 'link' if, by some procedure, we determine that two records refer to the same unit (Statistics NZ, 2015). Blocking and linking variables are used in linking records.

Appropriate selection of blocking and linking variables, and setting the cut-off weight, are key inputs to specifying the linking process. We use blocking to efficiently compare two datasets by reducing the number of records to compare. Records with the same values of a blocking variable will be compared before linking is carried out. For example, if blocking on date of birth, two records will only be compared if their dates of birth are identical.

Records are compared using the linking variables. We compare the values of linking variables for a pair of records to see the level of agreement between them. The probabilistic linking stage uses date of birth, sex, nationality, and the concatenated first and middle names as blocking variables. The linking variables are name, nationality, and the first and middle names (table 3).

The blocking and linking variables we used for this process, including the chosen cut-off weight, are given in table 3.

Table 3

Data integration identities for border movements

Pass	Blocking variable	Linking variable	Cut-off weight
1	Movement record identifier	…	0
2	Date of birth Sex	Name Nationality	21.25
3	Date of birth Sex of females Nationality First and middle names	First and middle names Name	21.26
Symbol: … not applicable Source: Stats NZ			

Linking passes

A 'linking pass' is an iteration of a record linkage process, using a particular set of blocking and linking variables. There are three passes in the linking process. The first pass matches all units with the same movement record identifier. The second pass uses a probabilistic match on name and nationality blocking on birth date and sex. The third pass matches on the first and middle name, and the whole name for females (to pick up possible name changes), and by blocking on birth date, sex, nationality, and the concatenated first and middle name. We created a sex variable for this pass for females (ie all males have a null value for this variable). (See table 3.)

Cut-off weights

When comparing two records, QualityStage assigns a field weight – a measure that reflects how similar they are. The field weight is calculated from two probabilities, the m- and u- probabilities.

The m-probability is the probability that the fields agree, given that the record pair is a match. It reflects the reliability of the data and can be computed as 1 minus the probability that the fields disagree, given the record is a match. When no estimates of the m-probabilities are available, for most fields we use 0.9, as shown in table 4.

Table 4

Input probabilities for linking variables

Pass	Matching field	m prob	u prob
2	name	0.9	0.01
2	dol_mid_nationality_code	0.9	0.01
3	firstmiddle	0.9	0.01
3	name	0.5	0.01
Source: Stats NZ			

The u probability is the probability that the fields agree, given that the record pair is not a match. It reflects the likelihood of a chance agreement and it is usually estimated by the inverse of the number of different values a field can take. For example, the u probability for sex is estimated as .5 as sex takes two possible values.

The cut-off weight is the weight chosen so that all record pairs with that weight or greater are regarded as matches, and all other record pairs are non-matches. We estimated the cut-off value for each pass from the frequency count of weights of the links created by QualityStage. After exploring different cut-off values and the matching results we determined approximate values for each linking pass. (Figure 6 illustrates an example, and table 3 gives the chosen cut-off values.)

Figure 7

Frequency count of probabilistic linking weight

Source: Stats NZ
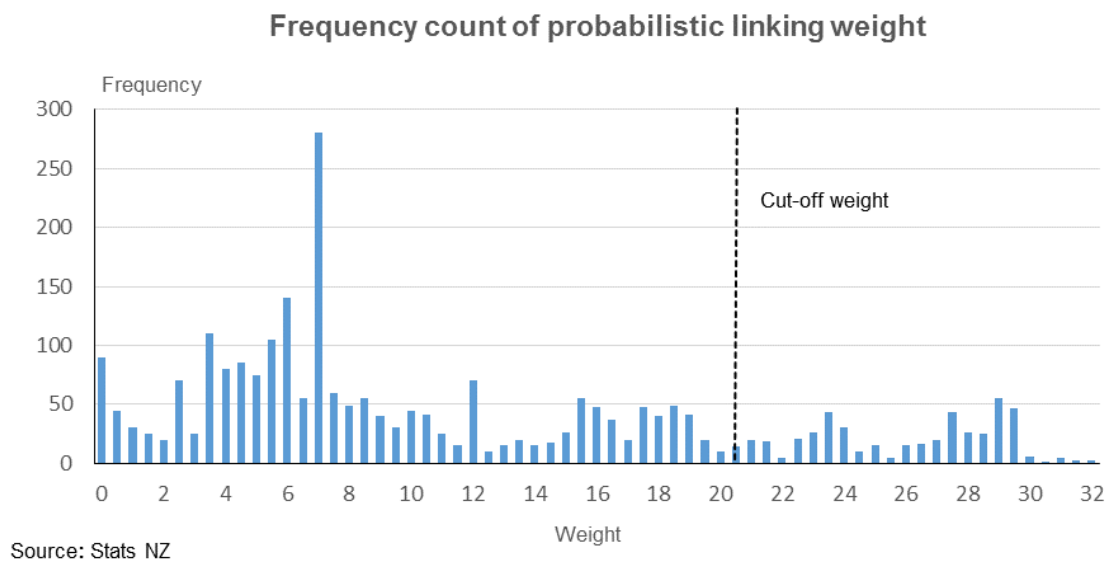
Appendix 3: Data integration results

This appendix provides an overview of the data quality assessment resulting from the data integration process of border movements described in appendix 2.

Linking results

There were 28 million distinct travellers in the links created. The link rate is calculated by counting the number of travellers who have linked (ie more than one movement entry) and dividing this by the total number of travellers within the reference period. We assume that all travellers should link although this may not be strictly true. People may not be travelling, and therefore have no movement recorded during that period, or they may have moved permanently overseas. The link rate was 98 percent.

Potential reasons for failing to link the remaining 2 percent include, but are not limited to, quality issues in the data, such as:

- females with changing surnames
- one person with two or more different birth dates
- a passport number was recorded incorrectly.

Linking quality

Quality can be measured by looking at the false positive rates of the linked data. False positives are the incorrect links picked up by the linking process. They are record pairs erroneously deemed to be links despite being true non-matches.

Generally, there is no good method for automatically estimating error rates, so false positive rates are estimated by manually checking samples of linked records. In large datasets, this estimation approach reduces the time and resources needed.

To estimate the false positive rate, we drew a sample of 200 travellers from each pass along with all the movement entries associated with them. This sample contained names (surname, and first names), birth date, sex, passport number, and all the movement information associated with a person – such as the movement carrier dates (date of travel) and the movement indicator (A - arrival or D - departure). Overall, we checked a total of 10,400 movement entries to identify the incorrect links.

The false positive rate for the data integration of border movements was 2.4 percent and the sampling error was 0.27 percent. This is comparable to other Stats NZ main data integration projects, such as the IDI's false positive rate of 2 percent.