

Population and migration statistics from administrative data

Methodology and evaluation report

Introduction

The census is carried out every 10 years and gives us a full picture of the size and makeup of the resident population in Jersey. Annual estimates of the size of Jersey's population and net migration have been produced by Statistics Jersey since 2003. These were produced using aggregated employment information together with estimates of migration behaviours. The estimates of migration behaviours were based on census information and were recalculated every 10 years. As such, they reflected the patterns of migration in the decade prior to the most recent census.

The population estimates were sufficiently robust in the years just after a census, and where there was stability in migration patterns. However, during periods of considerable change, such as during Brexit (the UK leaving the EU) and the Covid-19 pandemic, using parameters based on a historic point in time could not provide suitably robust estimates of the population.

An alternative way of producing population and migration statistics involves using administrative data (data that is already held by government for running public services). Administrative data is more current and wide-ranging and, when linked together, can provide evidence of people in Jersey accessing services. This can in turn determine whether someone can be classified as resident or not at points in time.

Some countries already use administrative data to produce population and migration statistics, for example Denmark, Finland, Guernsey, the Netherlands, Norway, Sweden, and Switzerland. This report summarises the methods, benefits, and challenges of the administrative data method for population estimation in Jersey.

Summary

Overall, the administrative data method provides a valuable means of estimating the size and structure of Jersey's population and migration, including increased granularity and accuracy relative to previous estimation methods.

Evaluation against the census showed the administrative data method produced a total resident population number at the end of March 2021 that was 0.14% (approximately 140 residents) higher than the 2021 Census (which related to 21st March 2021). At a more granular level, 97% of the residents counted by the 2021 Census were identified as resident by the administrative data method.

The methodology can be improved further, particularly using new data sources and modelling techniques to increase accuracy and reduce lag, so that timely annual population updates could be regularly produced. The linked data could also be used in future to analyse the journeys of cohorts, for example inward migrants as they move through industries over time, or groups of school leavers from different educational backgrounds and whether this impacts on their employment journey.

The detailed population and migration statistics, and potential for analysis of cohort journeys, can be used to support robust evidence-based decision-making in government and more widely.

Contents

Introduction	1
Summary	1
Contents.....	2
Legal basis	3
Methodology	3
Data sources	3
Overview	3
Creation of the person spine	4
Determining residency.....	5
Determining migration	7
Evaluation of method for estimating population	7
Matching accuracy.....	7
Characteristics selection	7
Choice of model to determine residency	7
Accuracy of chosen model evaluated against 2021 Census	10
Overall accuracy by age and sex	11
Potential bias	13
Benefits of administrative data for population and migration statistics – findings.....	13
The administrative data method is more accurate than the previous method of population estimation	13
The administrative data method can provide quarterly information	14
The administrative data method can provide inward and outward flows, not just net migration.....	15
The administrative data method can provide age-sex breakdowns, and by housing and work status	15
The administrative data method uses data already held – low respondent burden and cost effective	16
Challenges of administrative data for population and migration statistics – findings.....	16
Data protection.....	17
Safeguards	17
Safe setting	17
Safe people	17
Safe projects	17
Safe data	18
Safe outputs.....	18
Further work.....	18

Legal basis

The Statistics and Census (Jersey) Law 2018 states that Statistics Jersey:

- **Must** use administrative data wherever possible to produce statistics, particularly for the purposes of taking a census,¹
- **Can** access administrative data from government departments to produce statistics, and,
- **Must** protect that data to make sure no individual identifiable data is shared outside of Statistics Jersey

Methodology

Data sources

The following administrative data sources were used.

Table 1: Sources used in producing population and migration statistics from administrative data for the period 2017 to 2021

Data source	Department
Social Security benefits and contributions	Customer and Local Services
ITIS payments	Revenue Jersey
Employer manpower returns	Customer and Local Services
HCS demographic and appointment data	Health and Community Services
Preschool health checks and immunisations	Public Health
School roll data	Children, Young People, Education and Skills
Highlands roll	Children, Young People, Education and Skills
Student Finance	Children, Young People, Education and Skills
Youth service attendance	Children, Young People, Education and Skills
2021 Census data	Statistics Jersey (for evaluation purposes)

Overview

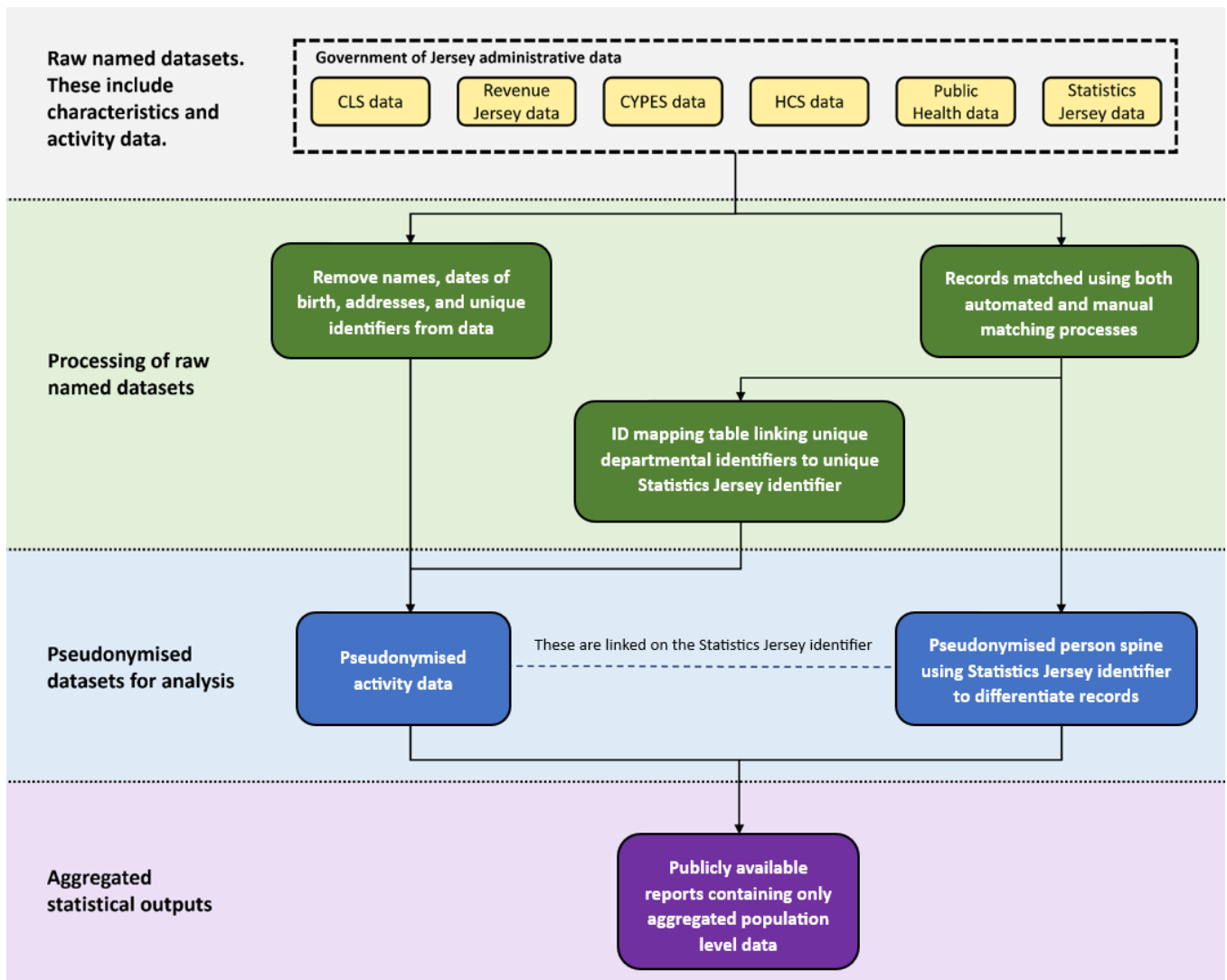
The administrative data is processed to create two different sets of data. The first is a pseudonymised dataset where each record represents one (and only one) person who has interacted with at least one government service since 2015. This is called the 'person spine'.

The second is a pseudonymised dataset of all the interactions people have had with the government services listed in Table 1 since 2015. This can be linked to the person spine through a unique Statistics Jersey identifier.

To achieve this, data is processed in a series of stages. Figure 1 briefly outlines the flow of administrative data through the different stages of data processing.

¹ [Statistics and Census \(Jersey\) Law 2018](#) Article 8, paragraph 2: 'for the purposes of a census, Statistics Jersey must, at such intervals as the Chief Statistician directs, analyse and link data collected for administrative purposes already in the possession of Statistics Jersey or readily obtainable by it.'

Figure 1: Process overview for administrative data project



Creation of the person spine

A number of different types of unique identifiers are used to identify people across the range of administrative datasets.

The process of identifying records across data sources that belong to the same individual is referred to as record linkage. Record linkage is performed by comparing the forename, surname, date of birth, and address fields of records across the data sources.

Record linkage for this project resulted in a single person spine – a list of the unique individuals that have been active in at least one administrative data source from the start of 2015 to present. Each unique person record was given a new unique Statistics Jersey identifier, and a mapping from this to the other unique identifiers used in the various administrative datasets was recorded. Only those administrative data sources with sufficient data quality were used to add new 'potential residents' to this person spine.

The automated matching process checked for:

- exact matches on fields
- phonetic matches on forename and surname (Soundex²)
- text based comparisons on combinations of forename, surname, and address that take account of spelling errors, abbreviations, and formatting (Levenshtein³ and Jaro-Winkler⁴ distances)

The automated matching process produces a score. Records with a high match score are assumed to belong to the same individual and automatically linked under the same unique Statistics Jersey identifier. Records with a medium match score were manually reviewed to determine whether they belonged to the same individual. Records with a low match score were automatically assumed not to belong to the same individual.

Certain person characteristics, such as sex and date of birth can be recorded differently in various data sources. In the population model, the age and sex used corresponds to the most frequently appearing values among the most recent records for that individual.

Determining residency

Any interactions that individuals have with services are referred to as 'activity'. Individuals are considered active in a quarter where they have particular patterns of activity in areas such as:

- Paying social security contributions
- Receiving social security benefits (including receipt of pension to a Jersey address)
- Attending GP appointments
- Receiving vaccinations
- Attending school or college
- Attending university
- Paying ITIS contributions
- Applying for and collecting a registration card
- Attending hospital appointments (and indicated to be resident at a point in time)
- In prison

Patterns in activity for individuals can be analysed to determine whether they were resident at a given time.

In the final model, a person is classified as resident at the end of a quarter if they have:

- activity in at least one quarter before that point in time, **and**
- activity in at least one quarter after that point in time, **and**
- a gap of less than 4 quarters between the two closest quarters of activity on either side of the point in time (see Figure 1).

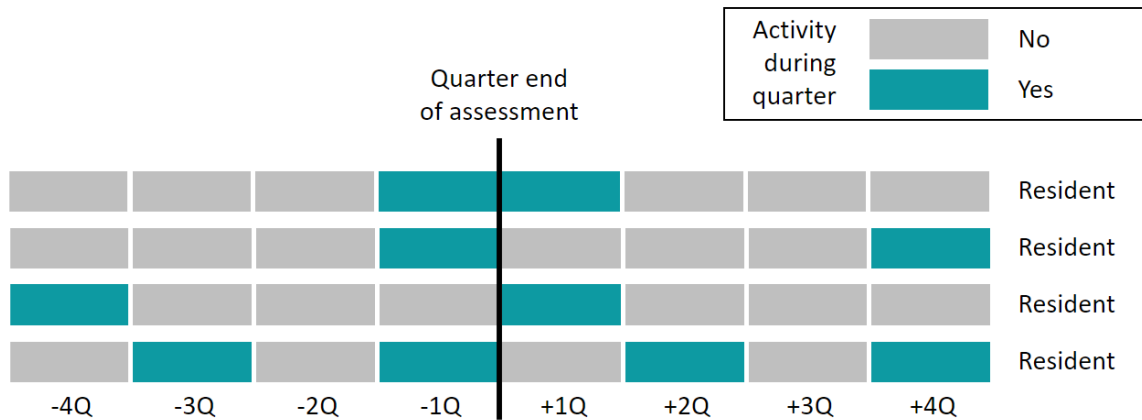
Under this method, it should be noted that seasonal workers will be counted as continuously resident if they appear in the activity data with gaps of three quarters or less, and if they have activity in Jersey during the same quarter in consecutive years – i.e. annually returning seasonal workers.

² Soundex is a phonetic algorithm for indexing names by sound, as pronounced in English.

³ Levenshtein distance is the minimum number of single-character edits (insertions, deletions, or substitutions) required to change one word into the one to which it is compared.

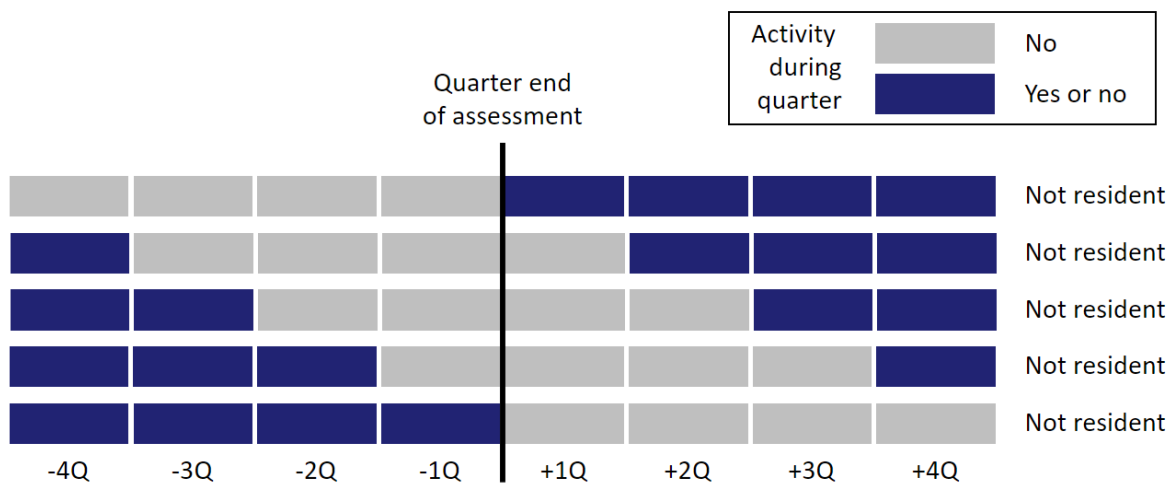
⁴ Jaro-Winkler distance measures the similarity between two sequences but places more emphasis on the beginning of the sequences. Where the beginnings of the sequences match, the score is more favourable.

Figure 1: Examples of activity patterns in which a person is defined to be resident in Jersey at a particular quarter end



In contrast, a person is classified as not resident if they have a period of four consecutive quarters with no activity directly preceding, directly following, or spanning the quarter end of assessment (see Figure 2).

Figure 2: The activity patterns in which a person is classified as not resident in Jersey at a particular quarter end



Determining migration

Whether someone in the person spine is classified as resident or not is determined at the end of each quarter. If they were determined as resident at the end of one quarter, but not at the one before it, they were assumed to have inwardly migrated during the period between the two points in time.⁵

If they were determined as not resident at the end of one quarter, but were resident at the one before it, they were assumed to have outwardly migrated between the two points in time.⁶

Under the administrative data method, if a resident left Jersey but returned within the subsequent four quarters, they would be counted as continuously resident, and would not be counted as an inward and outward migration.

Recurrent, short-term seasonal workers who return for the same one or two quarters every year would therefore count as continuously resident under this method.

People who come to Jersey for just one quarter and do not return in the subsequent four quarters would be excluded from migration patterns as they would not be deemed resident by the current model. This would apply to non-returning, short-term seasonal workers.

Evaluation of method for estimating population

Matching accuracy

The accuracy of the matching process was assessed by manually reviewing a random selection of 1,000 linked individuals. All of the records associated with the individuals from all the linked data sources were checked for erroneous links (a total of 6,277 unique link events). Of these, 8 links were found to be incorrect. This suggests that the accuracy of the matching process (i.e. the proportion of correct matches made by the matching process) is 99.9%.

Characteristics selection

For individuals who appeared in the 2021 Census, a comparison of the age and sex values selected by the model with the corresponding values in the census showed a 99.2% agreement rate for age and a 99.5% agreement rate for sex.

Choice of model to determine residency

Several different patterns of activity were trialled and evaluated for how well they assigned residency, by comparing the different outputs to the 2021 Census data.

Table 3 shows the different models and how well they performed when compared with the census data, against the following metrics:

⁵ The exception is when they had been born in that period.

⁶ The exception is when they had been recorded as having died in that period.

Overestimation – how many more or fewer residents the model identified compared to the census expressed as a percentage.

Precision – the proportion of people correctly classified as resident by the model over the total number of people classified as resident. This is given by the equation below.

$$\text{Precision} = \frac{\text{True positives}^7}{\text{True positives} + \text{False positives}^8}$$

Recall – the proportion of people correctly classified as resident by the model over the actual number of residents in the census dataset. This is given by the equation below.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}^9}$$

Year residence commenced (YRC)¹⁰ – a measure of the consistency of the year of most recent inward migration (or birth if continuously resident from birth) as calculated by the model when compared with the self-reported ‘year continuous residence began’ in the census.¹¹ This can only be calculated for those arriving between 2017 and 2021 (the period which the population model covers), and is not expected to reach 100% due to the different definitions and collection methods of the characteristic in the census and the administrative data model.

Time lag – how many quarters after a point in time can population numbers for that point in time be produced (this will include additional time required for the data quality to be at an optimal level, and processing time).

Generally, selecting the model with the highest precision and recall values implies fewer false positives and negatives. Models 1.01, 1.03 and 1.05 (see Table 3) all had precision and recall values at or above 97%. The lag for model 1.05 was considered too great and so this model was discounted. Although 1.03 had a better lag than 1.01, for the purposes of estimating the population 2011 to 2021, the difference between 4.5 and 5.5 quarters in the lag was not relevant.

Model 1.01 was selected to estimate the population for the period 2011 to 2021 as it showed better performance at estimating the overall size of the population. Also, the year residency commenced that was determined by this model more closely matched that seen in the census (although it should be reiterated that the different definitions and data collection methods for this characteristic mean that an exact 100% match between the model and the census would be an unreasonable target).

Model 1.01 classifies more people as resident than were recorded in the census data by 0.14% (approximately 140 residents). The amount by which a model overestimates or underestimates the number of census residents would ideally be a minimum. Of the models trialled, only one (model 1.10) had a smaller difference to the census total population number, though this model did not perform as well when considering precision and recall.

⁷ A true positive is where the population model determined a person was resident at the end of Q1 2021 and the person also appeared in the 2021 Census.

⁸ A false positive is where the population model determined a person was resident at the end of Q1 2021 but the person did not appear in the 2021 Census.

⁹ A false negative is where the population model determined a person was not resident at the end of Q1 2021 but the person did appear in the 2021 Census.

¹⁰ Year residence commenced (YRC) was a field recorded in the 2021 Census and is the year that an individual’s most recent period of residency began in Jersey.

¹¹ The question in the 2021 Census was: ‘When did your present period of continuous residence in Jersey begin? Ignore periods of absence on holiday and absences during the Occupation years.’

Table 3: Assessment of the models applied to determining an individual's residency when compared with 2021 Census data. The coloured lines are for comparison with other values in the column with green being relatively 'good' and red being relatively 'poor'.

Model	Pattern of activity 'not resident' (all other patterns were 'resident')	Overestimation (%)	Precision (%)	Recall (%)	YRC (%)	Time lag (Quarters)
1.00	4 or more consecutive quarters with no activity in the 4 preceding and 3 following quarters of date	0.83	96.8	97.7	70.2	4.5
1.01	4 or more consecutive quarters with no activity in the 4 preceding and 4 following quarters of date	0.14	97.1	97.4	70.5	5.5
1.02	3 or more consecutive quarters with no activity in the 3 preceding and 2 following quarters of date	0.35	96.9	97.4	65.2	3.5
1.03	3 or more consecutive quarters with no activity in the 3 preceding and 3 following quarters of date	-0.44	97.3	97.0	65.6	4.5
1.04	5 or more consecutive quarters with no activity in the 5 preceding and 4 following quarters of date	1.17	96.7	97.9	73.6	5.5
1.05	5 or more consecutive quarters with no activity in the 5 preceding and 5 following quarters of date	0.53	97.0	97.7	73.8	6.5
1.06	4 or fewer quarters with no activity in the 4 preceding and 4 following quarters of date	0.22	96.7	97.0	62.5	5.5
1.07	3 or fewer quarters with no activity in the 4 preceding and 4 following quarters of date	-1.99	97.5	95.7	64.3	5.5
1.08	3 or fewer quarters with no activity in the 3 preceding and 3 following quarters of date	0.57	96.5	97.2	60.2	4.5
1.09	2 or fewer quarters with no activity in the 3 preceding and 3 following quarters of date	-1.96	97.5	95.7	61.3	4.5
1.10	5 or fewer quarters with no activity in the 5 preceding and 5 following quarters of date	0.05	96.8	96.9	64.3	6.5
1.11	4 or fewer quarters with no activity in the 5 preceding and 5 following quarters of date	-1.91	97.5	95.8	66.8	6.5

Accuracy of chosen model evaluated against 2021 Census

The administrative data model for measuring the population can be evaluated against the results of the 2021 Census.

The two methods have different means of collecting the data (the census requiring every household and communal establishment in Jersey to record their residents and their characteristics through either an online or paper form), and different definitions of 'residency'. The census has traditionally been considered a 'gold standard' for statistics on population size and characteristics.

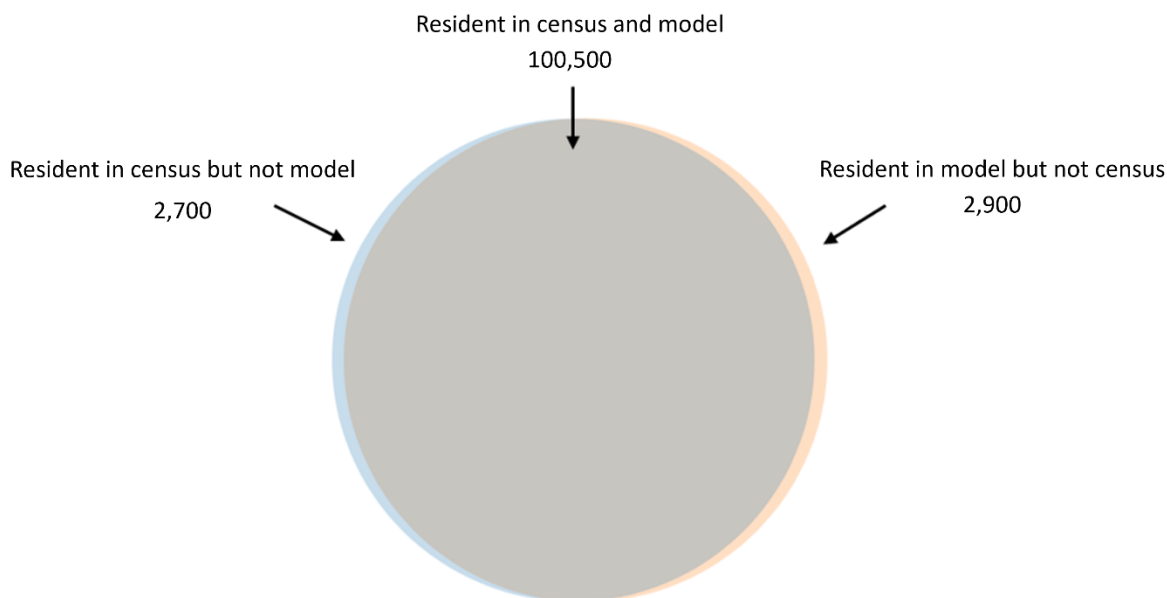
The census achieved a 99.5% response rate of households, and the missing persons were imputed,¹² giving a total of 103,267 persons at 21 March 2021. The administrative data model, which determined 103,400 to be resident at 31 March 2021, was compared to the full census dataset including the imputation for the nonresponse.

This is an overcount of approximately 140 residents (0.14%) when compared to the 2021 Census population figure.

It is important to note that in the 2021 Census, a person was considered resident 'if they lived or were intending to live in Jersey for one month or more'. This differs from the definition applied by the model (see section 'Determining Residency') which will result in different categorisations for some individuals.

Looking at an individual level, of the total number of residents in the model, 97.1% were also recorded as resident in the census. The proportion of total residents recorded in the census, who were correctly classed as resident by the model, was 97.4%. This is illustrated in Figure 3 which shows a Venn diagram of the residents in the model at the end of the first quarter in 2021 compared to those recorded as resident in the 2021 Census which took place on 21st March 2021.¹³

Figure 3: Venn diagram comparing people classified as resident by the model compared to residents recorded in the 2021 Census¹⁴



¹² Imputation is the process used to fill in missing records. For more information, please see the Statistics Jersey 2021 Census report.

¹³ The 'Resident in census but not model' group includes people who were imputed in the census. These people were not able to be linked to the model due to the nature of imputation. The 'Resident in model but not census' group may also include the people who were imputed in the census but could not be linked to their imputed census record.

¹⁴ All numbers have been independently rounded to the nearest 100.

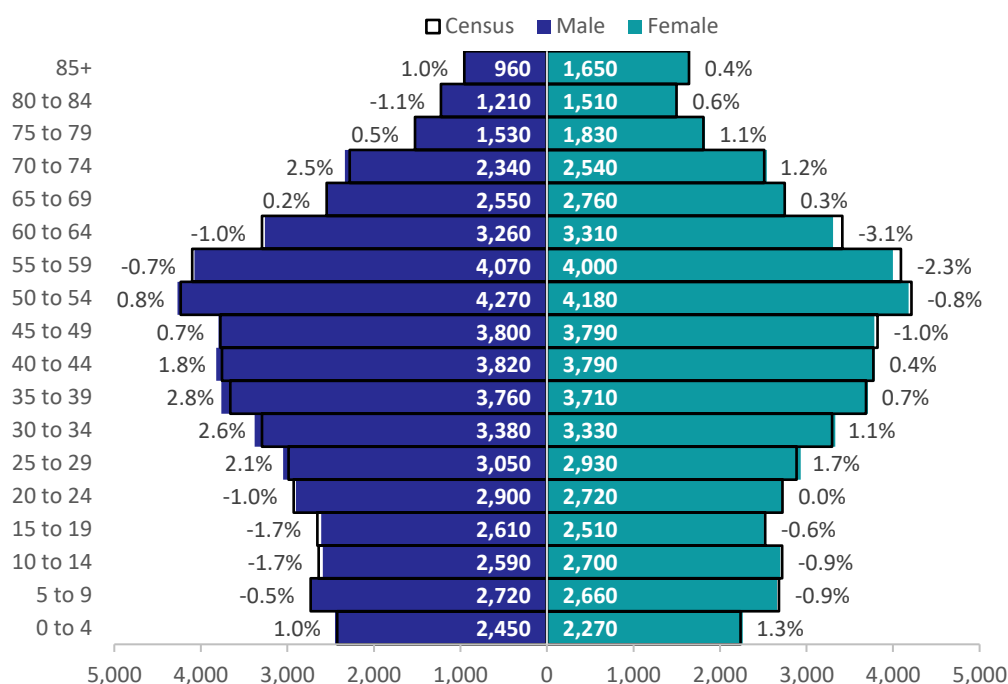
Overall accuracy by age and sex

The percentage value of the overcounts and undercounts of each age group and sex subpopulation are all within +/- 3.1% when compared to the 2021 Jersey Census.¹⁵ These percentages are shown in Figure 4 – overall there is strong similarity with the census population numbers.

For males in Jersey, the largest discrepancies in the subpopulation sizes were for 30 to 34-year-olds and 35 to 39-year-olds, where the administrative data method overcounted the population by 2.6% and 2.8% respectively, relative to the Jersey Census.

For females, the largest discrepancies were for 55 to 59-year-olds and 60 to 64-year-olds, where the administrative data method undercounted by 2.3% and 3.1% respectively, relative to the Jersey Census.

Figure 4: The population pyramid at the end of quarter one 2021, compared to the census, with the % difference between the model and the census for each cohort



Present in census but not in model, by age and sex

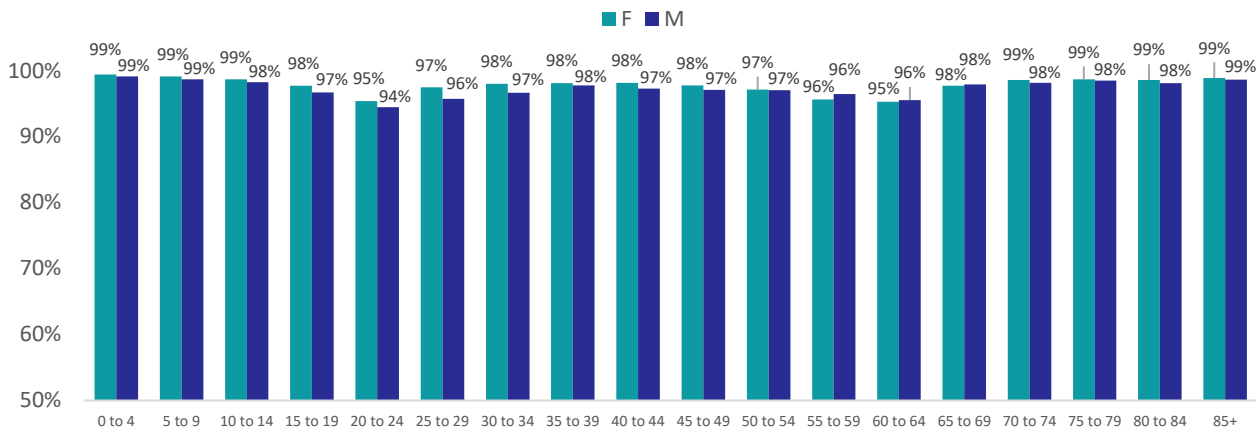
As shown in Figure 3, around 2,700 residents completed the census on 21 March 2021, but were not determined as resident at the end of March 2021 by the administrative data model. Further analysis of this group found the main contributors to be:

- Individuals with no or infrequent interactions with the government services included in the model (benefits, pensions, employment, health services, education)
- Self-employed individuals with infrequent interaction with government services
- Recent inward migrants not yet showing in administrative data
- University attendees not known to student finance
- Individuals working or travelling abroad

¹⁵ Although the ONS do not have defined quality standards for age group and sex subpopulation overcounts or undercounts, they do have a defined quality standard for Local Authority population size estimates in the UK. This standard requires all UK Local Authority subpopulation estimates to fall between +/- 3.8%, and that 97% of the subpopulation estimates fall between +/- 3.0%.

Figure 5 shows the proportion of people in the 2021 Census who were also determined as resident by the model by age and sex. The bars that are less than 100% show the subgroups where the model is not correctly classifying all the people resident in the census as resident. The model is more likely to incorrectly determine people in the 20 to 24 and 55 to 64 age ranges to be non-resident. The model also performs slightly worse for males with the exception of the 55 to 69 age range.

Figure 5: Proportion of people in the census who were also determined to be resident by the model by sex and age group



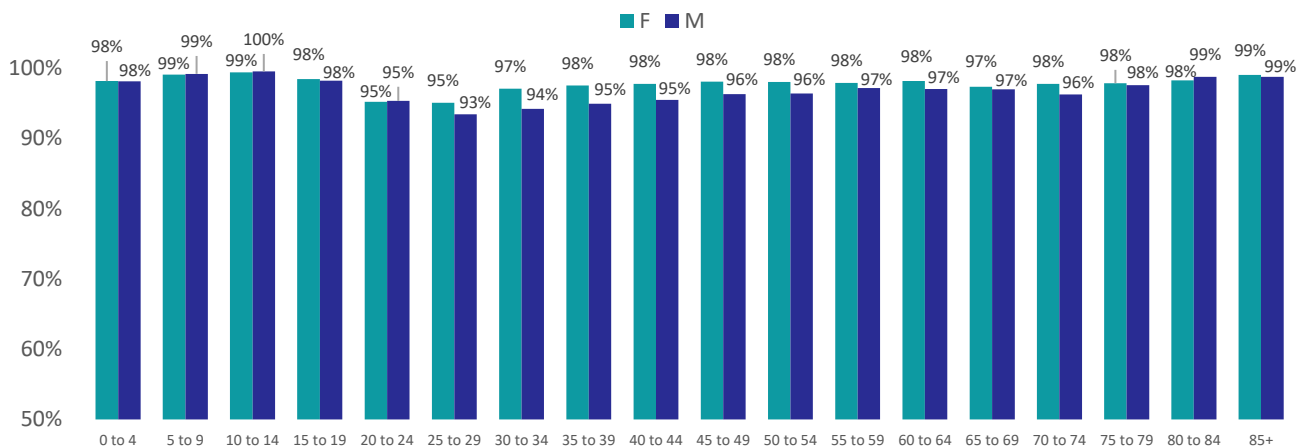
Present in model but not in census, by age and sex

There were around 2,900 people determined to be resident by the model at the end of the first quarter of 2021 that weren't recorded as resident in the census. Further analysis of this group found some of the main contributors to be:

- Differences in the model and census definition of residents
- Data quality issues in administrative data
- Undercount in the census¹⁶
- Seasonal workers with a gap in their employment at census time

It should be noted that in a number of cases there wasn't sufficient information available in the administrative data to be able to adequately determine a reason for the inclusion in the model and not the census. For example, where an individual was working remotely for a Jersey company, they would still show as resident in Jersey in the administrative data model but could have been living abroad and would therefore not be in the Jersey Census.

Figure 6: Proportion of people correctly classified as resident by the model when compared with the census by age group and sex



¹⁶ See the [Report on the 2021 Jersey Census](#) for information on the undercount.

The proportion of people correctly classified as resident by the model when compared with the census by age group and sex can be seen in Figure 6. The bars less than 100% show where people were classified by the model as resident but who weren't resident in the census dataset. This is greatest for the 20 to 29 age range and for all working age males in general.

Potential bias

As illustrated previously (see section '*Present in census but not in model*'), individuals who are resident in Jersey but have rare or no interaction with the administrative data sources collected will not be represented in the population statistics produced using this model. An example of this would be unemployed and healthy individuals who are not registered with Customer and Local Services as actively seeking work or in receipt of benefits. Another example would be older married women who still contribute and receive benefits via their husband's social security number and do not regularly interact with health services.

There is also potential for bias in the effectiveness of the matching algorithm used to link data. The function Soundex for example, rates like sounding names based on their sound in English and so may not be effective with non-English names. Further to this, the manner in which non-English names are recorded in administrative data sources leads to difficulty in linking them to the same individual. It is not uncommon for some naming conventions to use many forenames and surnames and only certain combinations of these to be recorded in some sources. It is possible that some links are missed meaning these subgroups could be under-represented. It is also possible that individuals could be duplicated in the person spine with each instance having only part of the activity profile, again potentially impacting whether they are determined resident.

The matching evaluation described above, whereby a manual inspection of a random sample of 1,000 records showed 99.9% correct matching, gives reassurance that any bias introduced through the matching process is minimal.

Benefits of administrative data for population and migration statistics – findings

The administrative data method for population estimation is:

- more accurate than the previous annual population estimation method
- able to provide quarterly information for the first time
- able to provide inward and outward migration flows for the first time
- able to provide breakdowns (of population size and migration) by age, sex and residential status for the first time
- has the potential to enable longitudinal analysis of cohorts in the future

Each of these is discussed briefly below.

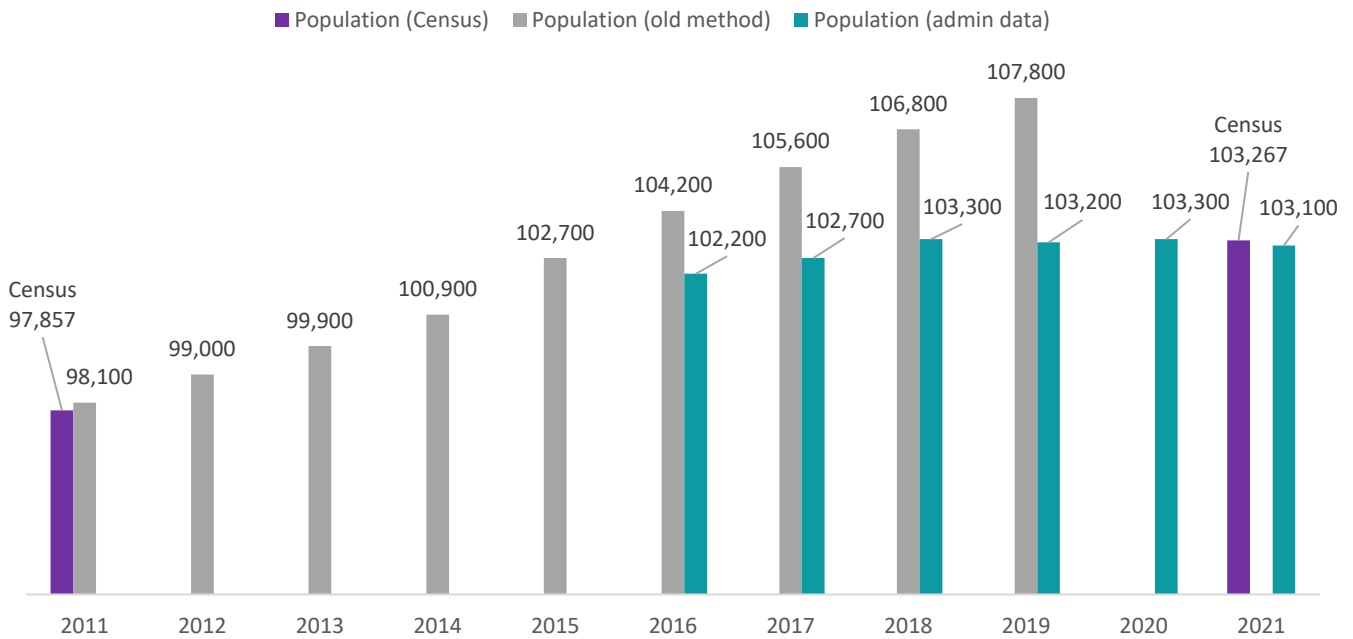
The administrative data method is more accurate than the previous method of population estimation

As Figure 7 shows, the administrative data methodology (which provides estimates from year end 2016) produces estimates of the population which align much more closely to that found in the 2021 Census.

The previous population estimate methodology used a cohort methodology. This took information from the 2011 Census, both to provide a baseline population and to provide estimates of the likelihood of inward migrants staying in Jersey over subsequent years (using the behaviour of migrants over the period 2001 to 2011). Each year following the census, known births and deaths and an estimate of net migration were applied to the 2011 baseline.

In contrast, the administrative data method does not use historic information to estimate current migration behaviour, and does not rely on making adjustments to a historic baseline, but each estimate is based on current data.

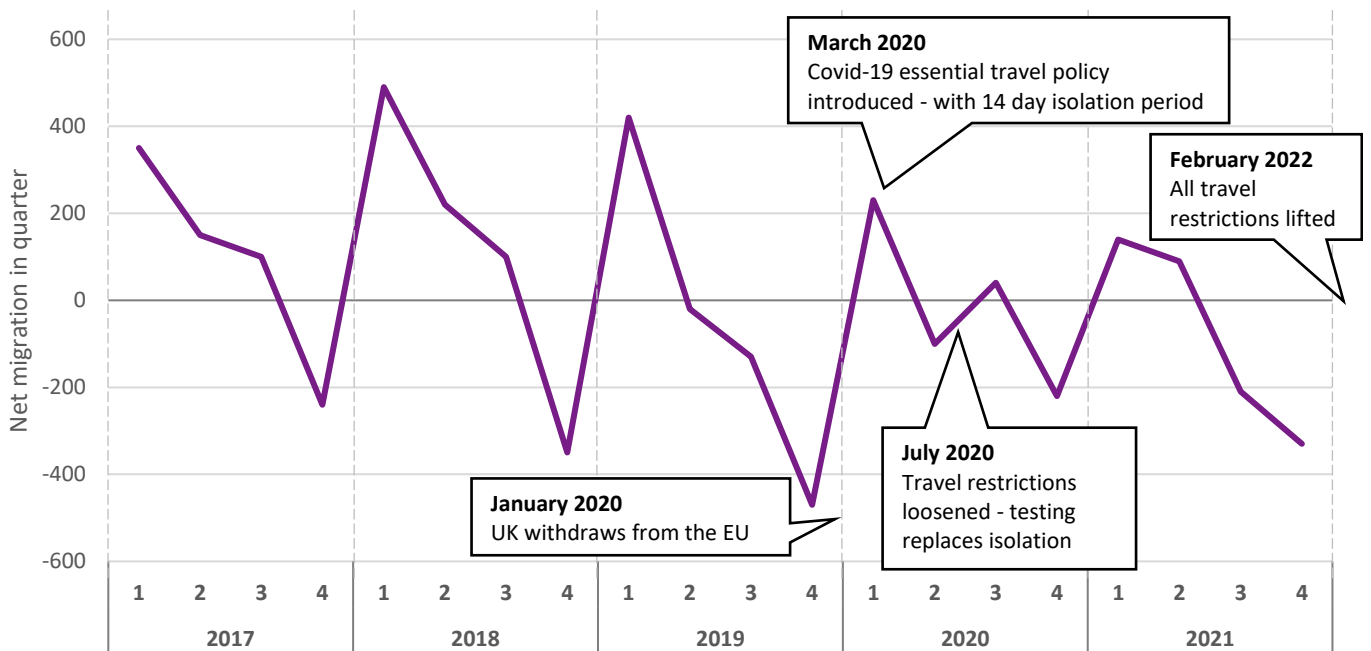
Figure 7: Population estimates as estimated by the previous methodology (at year end) and administrative data methodology (at year end) compared with 2011 and 2021 Census (March)



The administrative data method can provide quarterly information

The administrative data can produce more frequent estimates – currently quarterly estimates have been produced. These give a richer picture of the seasonal patterns of migration over time and are particularly helpful to look at how the decision for the UK to leave the EU and the pandemic may have disrupted these patterns.

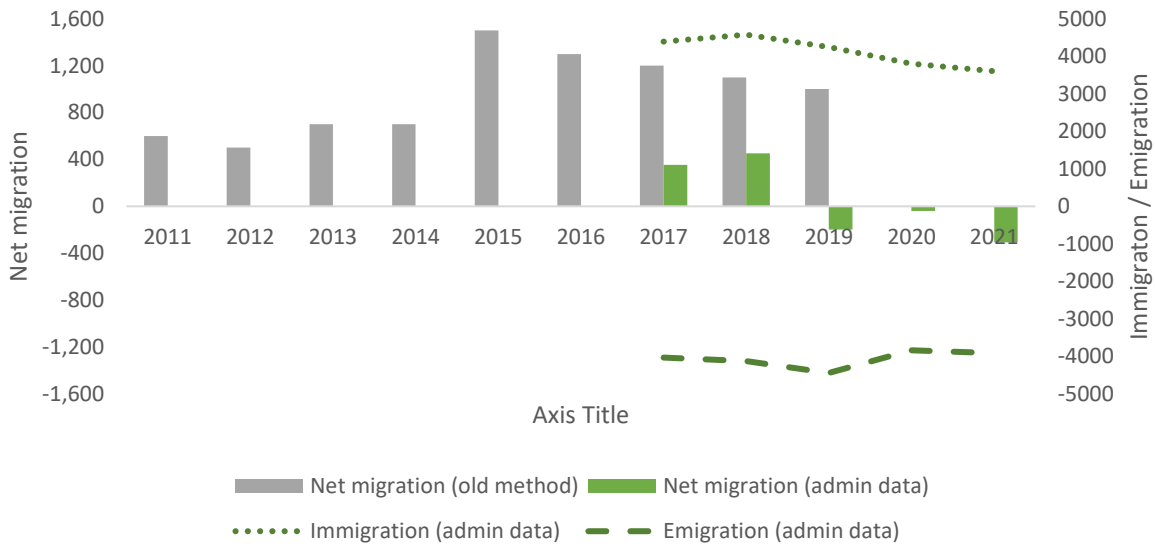
Figure 8: Quarterly net migration using administrative data methodology



The administrative data method can provide inward and outward flows, not just net migration

Previous estimate methodology only allowed estimation of net migration (the difference between the inward and outward migration flows) and not the size of the inward and outward flows themselves. The administrative data method works at an individual level and therefore can provide flow information.

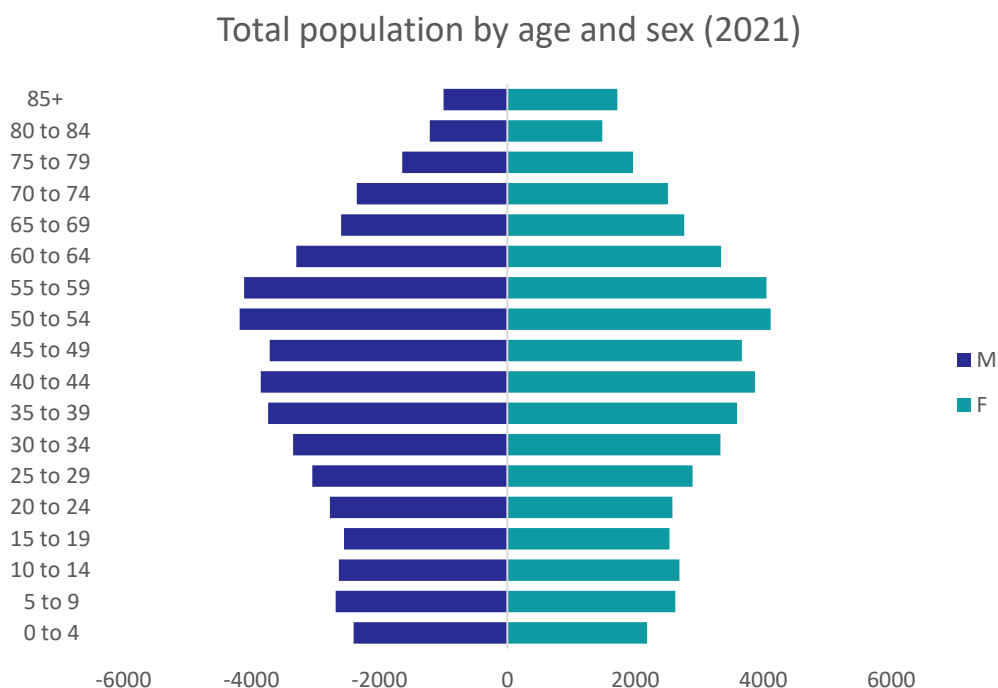
Figure 9: Net migration by year as estimated by previous methodology vs inward and outward migration flows and net migration as estimated by administrative data methodology



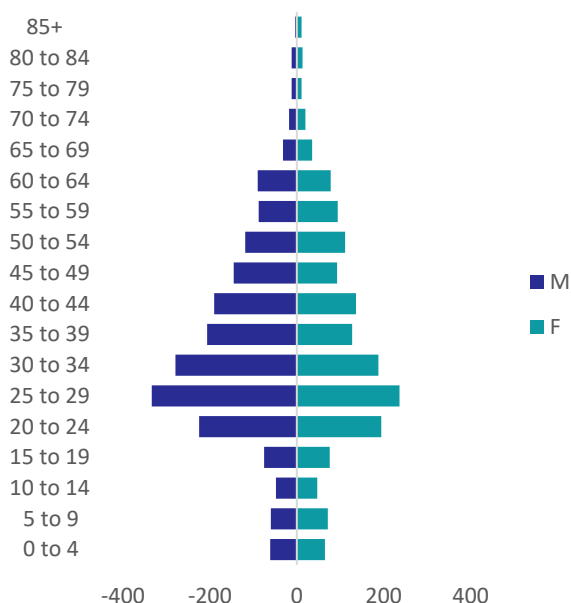
The administrative data method can provide age-sex breakdowns, and by housing and work status

The previous population estimates methodology was unable to provide an accurate age-sex breakdown of the population or migration. The administrative data method is based on individual pseudonymised records which have age and sex characteristics, and therefore breakdowns are possible. For those of working age, the model is also able to provide analysis by Entitled/ Entitled for Work, Registered or Licensed housing and work status.

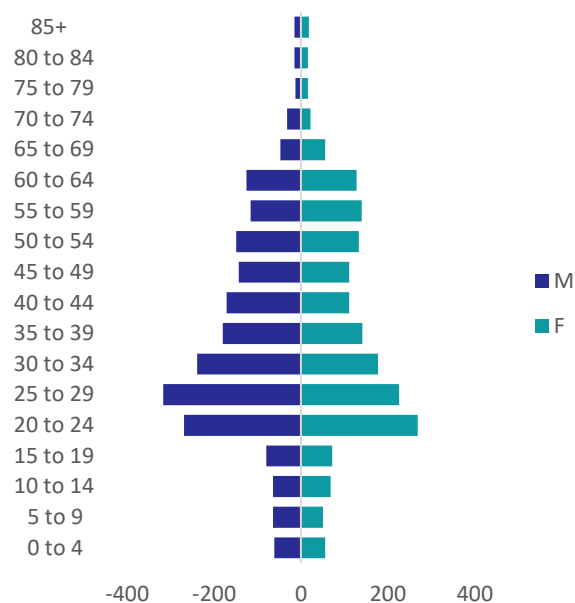
Figure 10: Age-sex breakdown of population, and migration flows, at each point in time is possible using administrative data methodology



Inward migration by age and sex (2021)



Outward migration by age and sex (2021)



The administrative data method uses data already held – low respondent burden and cost effective

Using data already held by government avoids needing to collect data from residents through a bespoke survey.

Challenges of administrative data for population and migration statistics – findings

- The data shared contains personal identifiable information, so that linking at individual level is possible between datasets. A Data Protection Impact Assessment was carried out, as was a review of the security measures in place around sharing, storing, and processing the personal data. Further details on the mitigation measures are provided in the 'Data protection' section.
- Administrative data is collected for administering public services and may not capture the required fields (for example ethnicity is not currently well captured in administrative systems). Linking a number of datasets together can help to reduce this issue.
- The data quality of administrative datasets is variable, and generally driven by the operational purposes that each dataset is serving. Combining datasets can help to reduce the impact of this on the statistics, though presents its own challenges. For example, poor data quality (such as inconsistency of formatting, lack of comprehensiveness) in key linking fields such as name, address, or date of birth increases the difficulty of matching across datasets and will increase the likelihood of mismatches being made. Manual inspection of 1,000 matched records in the data warehouse indicated that 99.9% of record links were matched correctly by the matching process.
- Some individuals who are resident will have minimal or no interaction with government services and may not be captured in the administrative data population and migration statistics.
- Not all data sources are produced regularly or with sufficient frequency.

Data protection

To link individual records across datasets, personal identifiable information needed to be shared with Statistics Jersey. A Data Protection Impact Assessment for producing population and migration statistics from administrative data was completed, along with Data Sharing Agreements with departments to ensure secure and documented processes and infrastructure for sharing and storing the data were identified, implemented and maintained.

Safeguards

A number of mitigations and safeguards have been put in place to protect the security and confidentiality of personal data. These have been detailed according to the 'Five Safes' approach,¹⁷ an approach used internationally by National Statistics Offices and administrative data research organisations.

Safe setting

The data is held on a dedicated, stand-alone server. Access to this environment is limited to specified employees within Statistics Jersey.

Within this environment is an even greater restricted area where the personal data is held. A minimal set of employees, working under the Statistics and Census (Jersey) Law 2018, have access to this area where the matching and linking takes place. Records are given a new Statistics Jersey unique identifier to enable records relating to the same people to be linked across datasets, and so that the personal data can be removed for the next stage of processing.

Still within the dedicated Statistics Jersey server, and only accessible to Statistics Jersey employees, the 'pseudonymised' data is held for analysis to be performed.

This is an internationally accepted approach by National Statistics Offices to safely and securely link administrative data.

Safe people

Only a subset of employees working for Statistics Jersey have access to the personal identifiable data. All employees working for Statistics Jersey have signed a declaration that they are aware of their legal obligations to not share or publish any individual identifiable data outside of Statistics Jersey. The declaration includes ensuring employees are aware of the penalties under the Statistics and Census (Jersey) Law 2018, if they were to misuse data, which include fines and imprisonment. Regular audits are taken of who has access to the personal identifiable data.

Safe projects

There is a legal basis for the project of producing population and migration statistics. Article 2, paragraph 3, of the Statistics and Census (Jersey) Law 2018 defines the functions of Statistics Jersey. These include:

- to collect, compile, analyse, extract, and disseminate data for statistical purposes
- to consider the statistical potential of the records maintained by public authorities and to facilitate the realising of this potential
- to conduct and report on censuses of the population of Jersey

¹⁷ [What is the Five Safes framework? — UK Data Service](#)

In addition, Article 12 of the Statistics and Census (Jersey) Law 2018 requires Statistics Jersey to:

- collect, whether in conjunction with any census or not, data relating to economic, social, demographic, and general activities and conditions in Jersey
- seek to avoid collecting data where data collected for administrative purposes is readily accessible, of suitable quality and sufficiently up to date to enable it to carry out its functions

Article 8, paragraph 2 of the Statistics and Census (Jersey) Law 2018 states that for the purposes of a census, Statistics Jersey must analyse and link data collected for administrative purposes already in the possession of Statistics Jersey or readily obtainable by it.

Safe data

As early as possible in the processing, the personal identifiable elements of the datasets are removed and replaced with a Statistics Jersey identifier. This enables analysis of pseudonymised data.

Safe outputs

In accordance with the Statistics and Census (Jersey) Law 2018, statistics produced using the linked administrative data will only ever be shared or published in aggregate form – in other words no individual identifiable data will be published or shared outside of Statistics Jersey.

Further work

Further work on the population model is planned to explore how to:

- reduce lag while maintaining accuracy
- improve accuracy of determination of residency through different activity patterns
- identify and integrate further data sources that could contribute to improving accuracy, particularly for those population groups currently less well captured by the model

In addition, the model itself has a wide application over and above population and migration statistics, such as:

- enabling longitudinal analysis of cohorts over time, for example, exploring how differences in educational settings are associated with different employment journeys, or the typical journey through employment (industries and occupation) of inward migrants
- modelling household characteristics and statistics, for example, deprivation by geography

Statistics Jersey

16 June 2023