

Statistical Research

Frequency: Ad hoc

Coverage: Northern Ireland



Date: 19 December 2024 (9.30am)

Geographical Area: Northern Ireland

Theme: Population

Time period: April 2021

Comparison of Northern Ireland Statistical Population Dataset with Census 2021

This research explores the creation of population estimates for Northern Ireland using a range of administrative datasets. It describes how these Statistical Population Dataset (SPD) estimates were created, what they estimate the Northern Ireland population to be in April 2021, and how they compare to the published census statistics (dated March 2021).

Contents

1	Key points	3
2	Introduction	4
3	Comparison of Northern Ireland Statistical Population Dataset with Census 2021	7
3.1	Comparison by sex	7
3.2	Comparison by age	8
3.3	Comparison by age and sex	9
3.4	Comparison by Local Government District	12
3.5	Comparison by Super Data Zone	16
4	Background Notes	19
4.1	Disclaimer.....	19
4.2	Statistical Research	19
4.3	Census 2021	20
4.4	Statistical Disclosure Control	21
4.5	Mid-Year Population Estimates	22
4.6	Data Acquisition	22
4.7	Data Processing.....	23
4.8	Usually Resident Population	24
4.9	Creating the SPD	25
4.10	Limitations.....	27
4.11	Future Developments	30
4.12	Link to Data used in this Research	31
5	Enquiries and Suggestions	32
6	Glossary	33

1 Key points

- The SPD estimated the population of Northern Ireland (NI) in April 2021 to be 1,897,600. This was slightly less (0.3%) than the official census population of 1,903,200.
- When split by sex, the SPD undercounted the census by 0.3% for both females and males at the NI level.
- The differences between the SPD and census were small for most ages. The SPD estimate was within $\pm 2\%$ for 71 out of 91 of the single years of age, and within $\pm 5\%$ for all single years of age.
- The SPD estimates by single year of age and sex were mostly within $\pm 2\%$ of census. The main exceptions were females aged 16 to 25 and males aged 19 to 27 years.
- Of the 11 Local Government Districts (LGDs), seven had an SPD estimate within $\pm 2\%$ of the census population. All LGDs had an SPD estimate within $\pm 5\%$ of the census. Of the four LGDs for which SPD and census populations differed by more than 2%, three had SPD undercounts and one had an overcount.
- Super Data Zones (SDZs) are a statistical output geography created to support the release of local-level Census 2021 statistics. There are 850 SDZs and almost half (46%) had an SPD estimate within $\pm 2\%$ of census. This increased to around three quarters of SDZs within $\pm 5\%$. Nine out of ten SDZs had an SPD estimate within $\pm 10\%$ of census.
- NISRA's research into the SPD supports future developments in official population estimates, since more timely population level administrative data can enhance the existing mid-year population estimates processes. The SPD estimates do not replace the accredited official statistics: the Northern Ireland [mid-year population estimates](#).

2 Introduction

NISRA used a range of administrative data sources to develop the Statistical Population Dataset (SPD). The aim of the SPD was to approximate the usually resident population by age and sex at the Northern Ireland (NI) and sub-NI geographic levels. The administrative data sources used in this research are maintained by different public bodies for operational purposes, such as the provision of services. Using administrative data for secondary purposes is good practice under the value pillar of the [Code of Practice for Statistics](#)¹.

Disclaimer: The Statistical Population Dataset (SPD) estimates do not replace the accredited official statistics: the Northern Ireland mid-year population estimates². The SPD estimates use a different methodology to the NISRA mid-year population estimates. Any use of these outputs should make clear to users the nature and purpose of the research and should include this disclaimer. For more information see the [disclaimer section](#). NISRA consider these outputs to be [statistical research](#) rather than official statistics in development.

NISRA's research into the SPD supports future developments in official population estimates. NISRA's mid-year population estimates are produced using a range of administrative data sources and are compared to several administrative data sources to assure the quality of outputs. More timely population level administrative data as provided by the SPD can enhance the existing mid-year population estimates processes by providing a population estimate created using a different methodology, giving accurate results (with some limitations, as discussed in this report) and with the ability to produce estimates at a reference point close to the mid-year population estimates (30 June). Furthermore, the SPD estimates in this paper are aggregates from an individual record level dataset. In contrast the mid-year population estimates methodology starts with aggregated data. One of the advantages of aggregating from an individual record level dataset is greater flexibility when choosing which geographic areas to aggregate to e.g. statistical, administrative or even bespoke geographies, subject to quality and disclosure control.

¹ Principle 4 (innovation and improvement) and principle 5 (efficiency and proportionality).

² NISRA accredited official statistics for population are the [mid-year population estimates](#).

This work builds on the successful use of administrative data in the 2011 and 2021 censuses and on previous research in the use of administrative data in population estimates. Further information on the uses of administrative data is available on the NISRA website:

- [Use of administrative data in population estimates](#)
- [Data Matching Using NI Administrative Data: A Worked Example](#) (pdf, 650 KB)
- [NI Census 2011 general report](#)³
- [Census 2021 use of administrative data](#) (pdf, 251 KB)
- [Census 2021 methodology overview](#) (HTML)
- [Census 2021 quality assurance report](#) (HTML)⁴

In October 2014, NISRA published the use of administrative data in population estimates paper, which reported on a prototype population estimate constructed using aggregated administrative data. In its conclusion it described how an improved administrative data based population estimate could be created through the introduction of further sources, ideally at the individual record level, and adjusted through a coverage survey.

This paper describes the results of several developments in the use of administrative data for population estimates. This includes advances in NISRA's data matching capabilities, combining additional linked administrative datasets at the individual record level and developments in how the combined datasets were used to best estimate the population. It demonstrates how NISRA's SPD research complies with the quality pillar of the [Code of Practice for Statistics](#). Principle 3 (assured quality) in particular, and also principle 1 (suitable data sources) and principle 2 (sound methods). This paper investigates the quality of the SPD by comparing different aggregations of age and sex, at different geographic levels, to published NI Census 2021 figures to better understand where and why it overcounted or undercounted the population. The comprehensive nature of the census makes it crucial in a wide range of different settings. It is used as a baseline for the creation of annual population estimates and, due to the detailed picture that it provides of the population of Northern Ireland, provides a high quality benchmark for this research. This SPD has been created at April 2021 to enable comparison with the census. NISRA intend that future research will

³ The coverage assessment and adjustment section included the use of administrative data.

⁴ Administrative data was mentioned in several sections of this quality assurance report.

explore this further by analysing individual record level data. See the [future developments](#) section of this report for more information.

In this report the term “census” refers to NI Census 2021. For display purposes, figures are appropriately rounded (to the nearest 100 for counts and 0.1% for percentages). Unrounded figures are available in the tables that accompany this report.

3 Comparison of Northern Ireland Statistical Population Dataset with Census 2021

3.1 Comparison by sex

The population of Northern Ireland (NI) in April 2021 was estimated to be 1,897,600 according to the Statistical Population Dataset (SPD). This was slightly less (0.3%) than the official census population of 1,903,200 on census day, 21 March 2021. The 95% confidence interval for the census population was $\pm 0.25\%$ ⁵.

Table 1: Northern Ireland population from SPD and census, 2021

Sex	SPD	Census	Difference	Percentage difference
Female	964,600	967,000	-2,500	-0.3%
Male	933,000	936,100	-3,100	-0.3%
Northern Ireland	1,897,600	1,903,200	-5,600	-0.3%

[Download data](#) (xlsx, 55 KB)

Note: Figures may not sum due to rounding.

The SPD undercounted the census population by 0.3% for both females and males. The analysis that follows considers different aggregations of the SPD by combinations of age, sex, Local Government District (LGD) and Super Data Zone (SDZ) and compares each aggregation to the comparable statistics from the census.

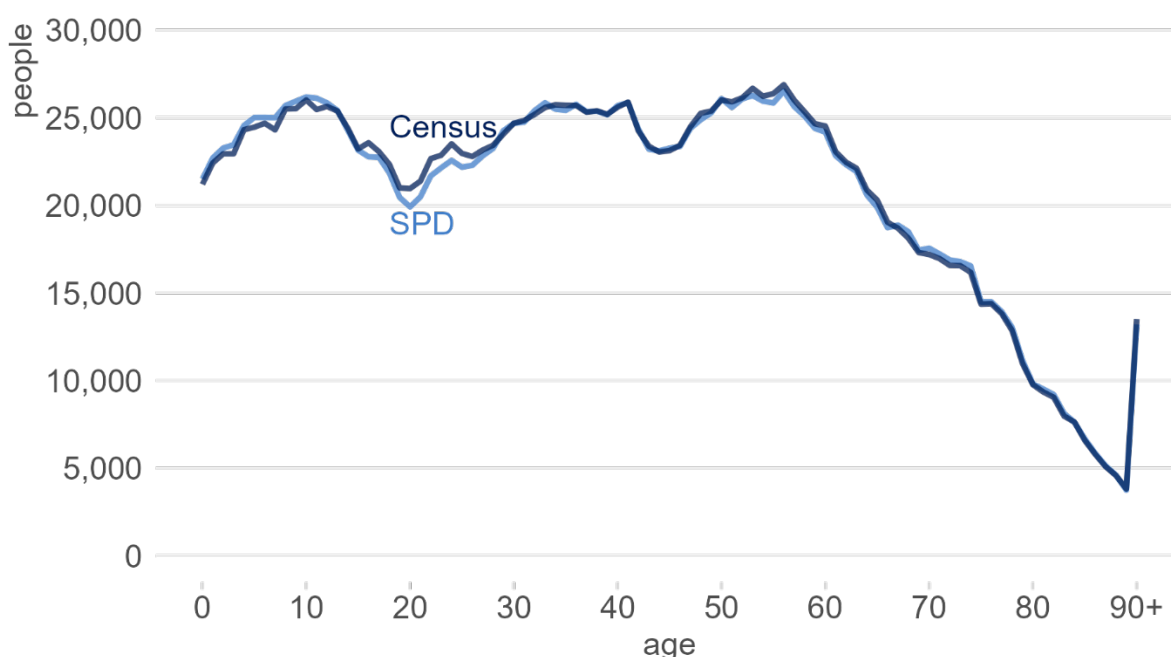
It is important to note that the SPD can both undercount and overcount different subsets of the census population. The administrative data used to create the SPD can contain people present in the census data; can contain people not present in the census data; and can also not contain (be missing) people present in the census data.

⁵ $\pm 4,700$ residents, from the Census 2021 [coverage information paper](#) (page 14).

3.2 Comparison by age

The differences between the SPD and census population estimates were small for most ages, as shown in Figure 1. The SPD was within $\pm 2\%$ of the census population for 71 out of 91 of the single years of age, and within $\pm 5\%$ for all single years of age. Note that the eldest age group combined all ages 90 years and older to prevent the disclosure of small population counts.

Figure 1: Population estimates by age from SPD and census, 2021



[Download Chart](#) (xlsx, 156 KB)

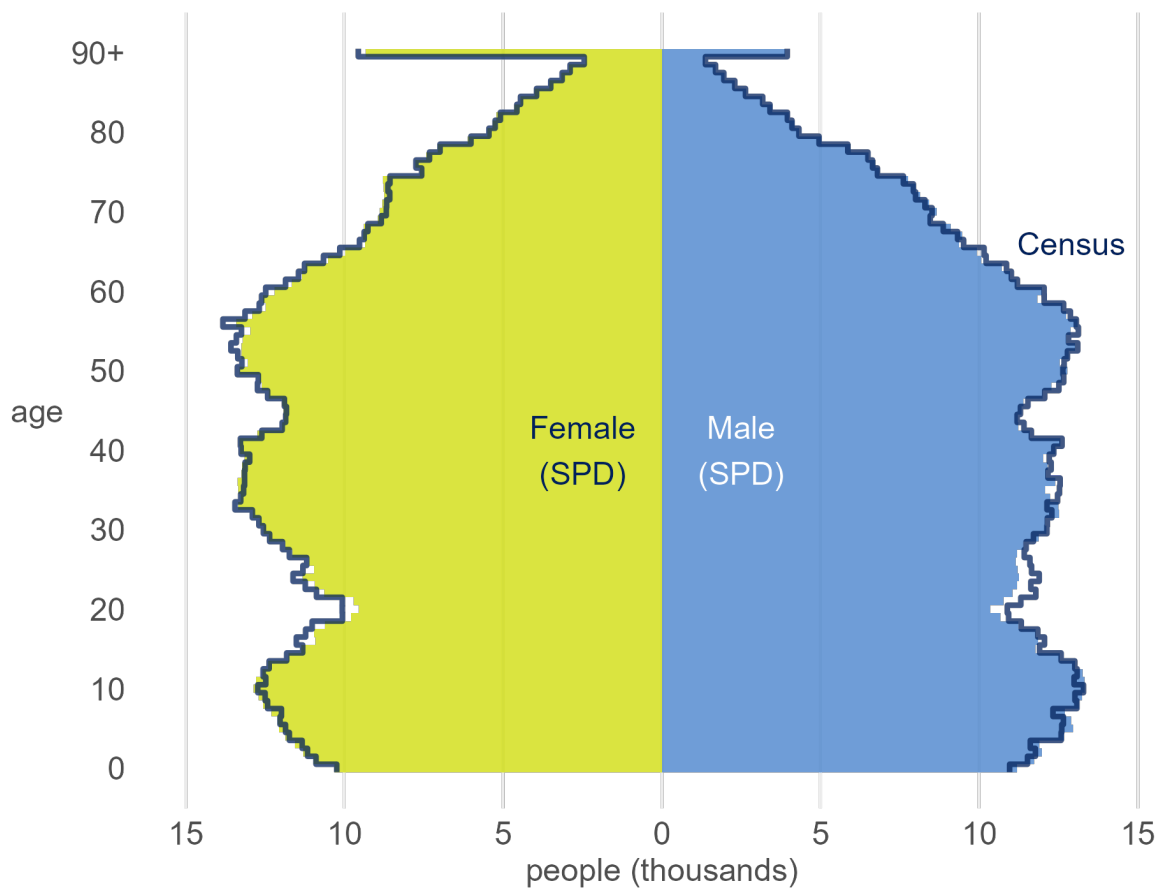
Note: '90+' represents 90 years and older.

The SPD estimates were generally higher than census for children up to 12 years of age and for adults aged 67 to 88 years old. These overcounts tended to be small with most of the affected ages having overcounts of less than 2%. The SPD estimates were less than census for people aged 16 to 28 years, 51 to 66 years and those older than 89 years of age. The greatest undercounts (less than -2%) were for young adults aged 16 to 26 years old, as can be seen in Figure 1.

3.3 Comparison by age and sex

Figure 2 compares two population pyramids by single year of age and sex. The horizontal bars are SPD estimates for females and males in April 2021 and the dark outline shows the Census 2021 population for comparison from March 2021. Note that the eldest age group combines all ages 90 years and older to prevent the disclosure of small population counts.

Figure 2: Population by age and sex from SPD and census, 2021

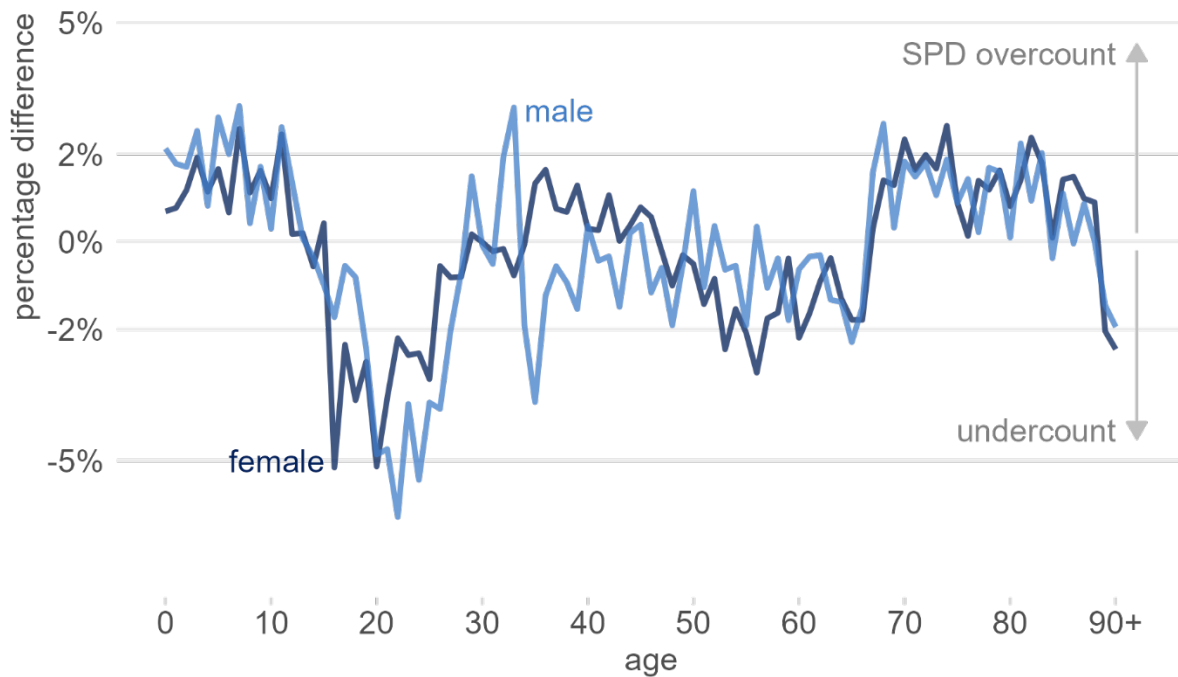


[Download Chart](#) (xlsx, 160 KB)

Note: '90+' represents 90 years and older.

The line chart in Figure 3 shows the percentage difference between the SPD estimates and census by age and sex. The differences were similar for both sexes showing SPD undercounts (percentage differences below 0%) and overcounts (above 0%) for similar age ranges. The SPD estimates by age and sex were mostly within $\pm 2\%$ of census. The most notable exceptions were the females aged 16 to 25 and males aged 19 to 27 years. For both these groups the SPD undercounted the census, which can be seen in Figures 2 and 3.

Figure 3: Percentage differences between SPD and census by age and sex, 2021



[Download Chart](#) (xlsx, 241 KB)

Note: '90+' represents 90 years and older.

The SPD estimates by age and sex were almost entirely within $\pm 5\%$ of census, as shown in Figure 3. This analysis considered 91 single years of age⁶, which gave 182 age and sex combinations. There were only four age and sex combinations for which the SPD estimate differed from census by more than 5%. These differences were undercounts and occurred for females aged 16 and 20 (-5.2% and -5.1% respectively), and males aged 22 and 24 (-6.3% and -5.4% respectively).

When overcounting was observed it tended to affect under 14 and over 66 year olds. This can be seen in Figure 2 and more clearly in Figure 3. Most single years in these age ranges still had SPD estimates within $\pm 2\%$ of census.

Although the SPD and census were broadly similar there were differences for which the SPD was too high (overcounting) and too low (undercounting). The SPD can contain people not present in the census data, causing overcounting. A person can appear in an administrative

⁶ As '90+' represents age 90 and over, to prevent the disclosure of small population counts.

dataset but then emigrate from NI. In some cases the person may continue to be counted in the SPD until they no longer meet the [activity-related inclusion rules](#) for the various administrative datasets that they had appeared on. Population sub-groups with greater levels of population churn (how much a population moves into and out of an area), including younger adults and student areas, may have overcounting due to the time lags involved with their details being updated on the administrative datasets.

The SPD can be missing people who were present in the census, causing undercounting. A person can live in NI but not appear on any of the available administrative datasets. A person who had recently moved to NI but had not yet interacted with one of the administrative datasets used to create the SPD would be missed. Alternatively, a person could appear in one or more of the administrative datasets but not recently enough to be included. It is possible that younger adults may interact less with the particular administrative data sources used to create the SPD.

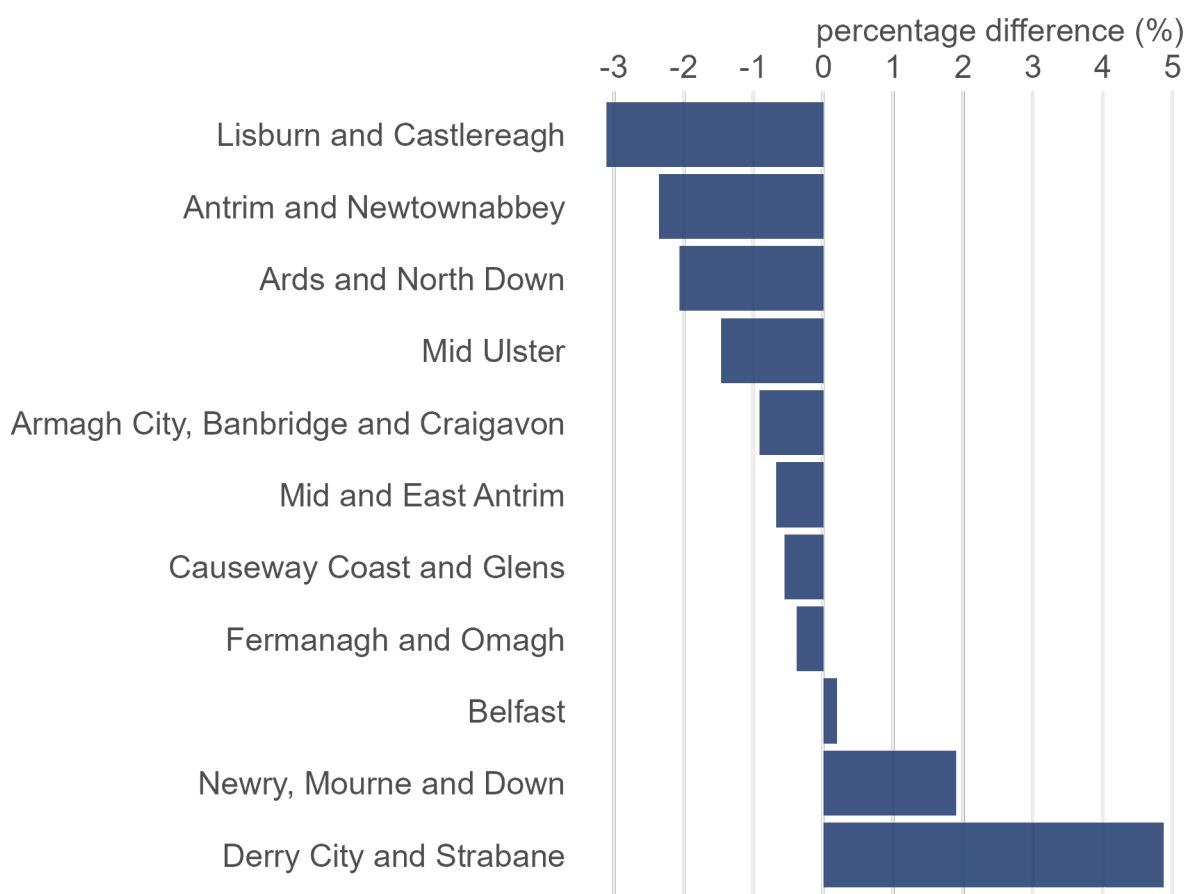
The SPD can contain people present in the census data but at different addresses. This could cause an overcount in one area and an undercount in another area. A person can also have different addresses in different administrative datasets, which the SPD methodology accounts for. If a person has moved address but has not yet updated their address on any of the administrative datasets, then the SPD would assign an inaccurate location.

For more information about the SPD methodology and limitations see the [creating the SPD](#) and the [limitations sections](#) of this report. The issues relating to undercounting and different addresses could be mitigated by incorporating additional administrative data into the SPD, in order to more accurately estimate the sub-populations that SPD has undercounted, like females aged 16 to 25 and males aged 19 to 27 years. For example administrative data relating to tax records and migration.

3.4 Comparison by Local Government District

There are 11 Local Government Districts (LGDs) in NI. LGDs provide a useful sub-national geographic level to analyse differences between SPD and census. Most LGDs (7 out of 11) had an SPD estimate within $\pm 2\%$ of the census population, as shown in Figures 4 and 5. All LGDs had an SPD estimate within $\pm 5\%$ of the census. Of the four LGDs for which SPD and census populations differed by more than $\pm 2\%$, three had undercounts and one had an overcount.

Figure 4: Percentage differences between SPD and census by LGD, 2021



[Download Chart](#) (xlsx, 180 KB)

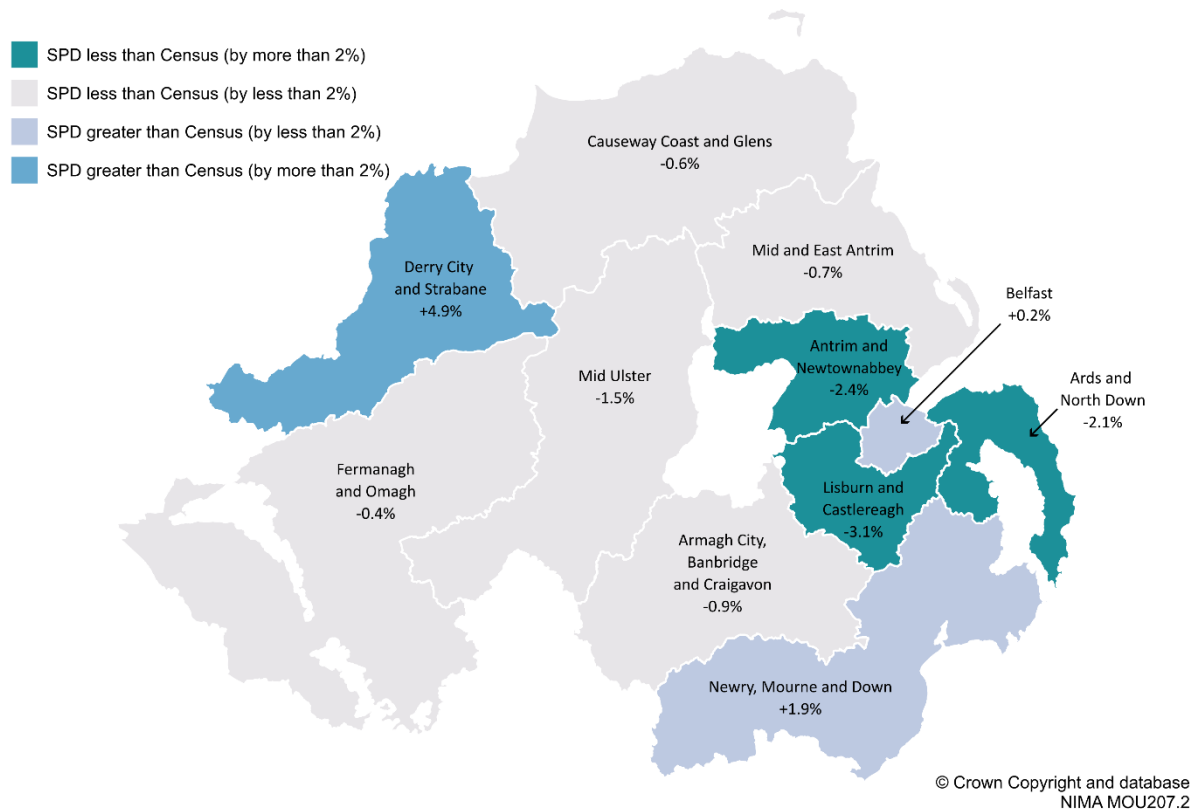
Note: percentages less than zero indicate an undercount (SPD less than Census 2021).

The four LGDs with percentage differences greater than 2% had some of the greatest proportions of usual residents with urban status⁷ in the census. Belfast LGD had the greatest

⁷ [Usual resident population by LGD and Urban Status CT0116](#)

proportion of its population with urban status (98.7%) but the next four highest LGDs were Antrim and Newtownabbey (77.3%), Ards and North Down (72.8%), Derry City and Strabane (64.6%) and Lisburn and Castlereagh (63.8%).

Figure 5: Map showing percentage difference between SPD and census by LGD, 2021



[Download Map and Data](#) (xlsx, 248 KB)

Note: percentages less than zero indicate an undercount (SPD less than Census 2021).

For each LGD, the SPD and census populations were analysed by sex and five year age band, as single year of age resulted in relatively small population counts.

3.4.1 Local Government District with greatest overcount

Most LGDs (8 out of 11) had an SPD estimate less than the census population as shown in Figures 4 and 5. The LGD with the greatest SPD overcount when compared to census was Derry City & Strabane LGD with 4.9% (7,300 people). In Derry City and Strabane, the SPD estimate exceeded census for every five year age and sex group except for the 90+ groups. The largest overcounts (over 5%) were for females aged 25-44 and males aged 25-54 years.

It is possible that the proximity of this LGD to the border between NI and Ireland may have resulted in a greater population churn and contributed towards the SPD exceeding the census. Also Derry City and Strabane was the LGD with the second highest proportion of usual residents above compulsory school age⁸ who were full-time students (6.8%)⁹ and was also the LGD with the second highest proportion of usual residents who lived in Ireland one year ago (0.3%) according to the census¹⁰. Central Statistics Office (CSO) Ireland's figures show that 2,400 Ireland residents were usually resident in NI in the year leading up to Census 2022¹¹. These features made it more difficult for the SPD to accurately estimate the usually resident population in areas close to the border between Ireland and Northern Ireland.

3.4.2 Local Government District with greatest undercount

The LGD with the greatest SPD undercount when compared to census was Lisburn and Castlereagh with 3.1% (4,600 people). This can be seen in Figures 4 and 5. In Lisburn and Castlereagh LGD, the five year age and sex groups with the largest undercounts (less than 5%) were females aged 25-29 and males aged 20-49.

It could be that the undercounts in these LGDs related to people who did not interact regularly with the administrative datasets used to create the SPD, for example, young working adults in good health. The census provides the proportion of usual residents that self-reported their general health as being very good, good or fair, and who were in

⁸ [Compulsory School Age](#) indicates whether a person was aged 16 or over on 1 July 2020.

⁹ Census 2021 Flexible Table Builder: [Compulsory School Age by Full Time Student by LGD](#)

¹⁰ Census 2021 Flexible Table Builder: [Address One Year Ago \(6 categories\) by LGD](#)

¹¹ [CSO Census 2022 Table FY019](#): Population Aged One Year and Over Usually Resident and Present in the State Who Lived Outside the State

employment¹². The LGD with the highest proportion was Lisburn and Castlereagh (47.5%). For comparison the overall figure for NI was 43.9%. Furthermore, some of the undercount in the Lisburn and Castlereagh LGD was particularly apparent in the geographical areas close to large communal establishments¹³.

3.4.3 Local Government District with both overcounts and undercounts

Belfast had an SPD estimate very close to the census population, exceeding it by just 0.2% (700 people). However, this apparent accuracy concealed both overcounting and undercounting issues for different five year age and sex groups. The largest undercounts (less than -5%) were the 20-29 year olds for both sexes and in particular the 20-24 year old males (-12.7%). The largest overcounts (more than 5%) were males aged 40-44 years old.

There may be multiple reasons for the overcounting and undercounting in the Belfast LGD. A more mobile population, relative to the rest of NI, could result in time lags in the administrative data, resulting in people being counted in the wrong location. In the census Belfast LGD had a greater proportion of usual residents who lived at a different address one year ago (12.5%) than the next highest LGD (Ards and North Down, 8.4%) by some margin.¹⁴

A large student population could also contribute to population churn. Belfast LGD had the highest proportion of usual residents aged above compulsory school age⁸ who were full-time students (10.9%)¹⁵. This figure was considerably higher than the equivalent NI figure (6.9%).

The [limitations section](#) contains more information about how the SPD estimates can overcount and undercount the census population. More administrative data concerning recent migration could help to resolve these issues.

¹² Census 2021 Flexible Table Builder: [Economic Activity by Health in General by LGD](#)

¹³ In Census 2021 a [communal establishment](#) is an establishment providing managed residential accommodation. For example, care homes and education establishments.

¹⁴ Main statistics for NI: [Statistical bulletin - Migration](#)

¹⁵ Census 2021 Flexible Table Builder: [Compulsory School Age by Full Time Student by LGD](#)

3.5 Comparison by Super Data Zone

Super Data Zones (SDZs) are a statistical output geography created by NISRA to support the release of local-level Census 2021 statistics¹⁶. SDZs nest within Local Government Districts. There are 850 SDZs with an average population of 2,200 usual residents (Census 2021). The smallest SDZ contains 1,100 usual residents and the largest contains 4,100 according to the census. Their small and relatively stable populations make SDZs a useful geographic level on which to compare local-level differences between SPD and census.

Table 2 shows that almost half (45.6%) of SDZs had an SPD estimate within $\pm 2\%$ of census. This increased to around three quarters of SDZs within $\pm 5\%$. Nine out of ten SDZs had SPD within $\pm 10\%$ of census.

Table 2: Number and percentage of SDZs with SPD within 2, 5 or 10% of census

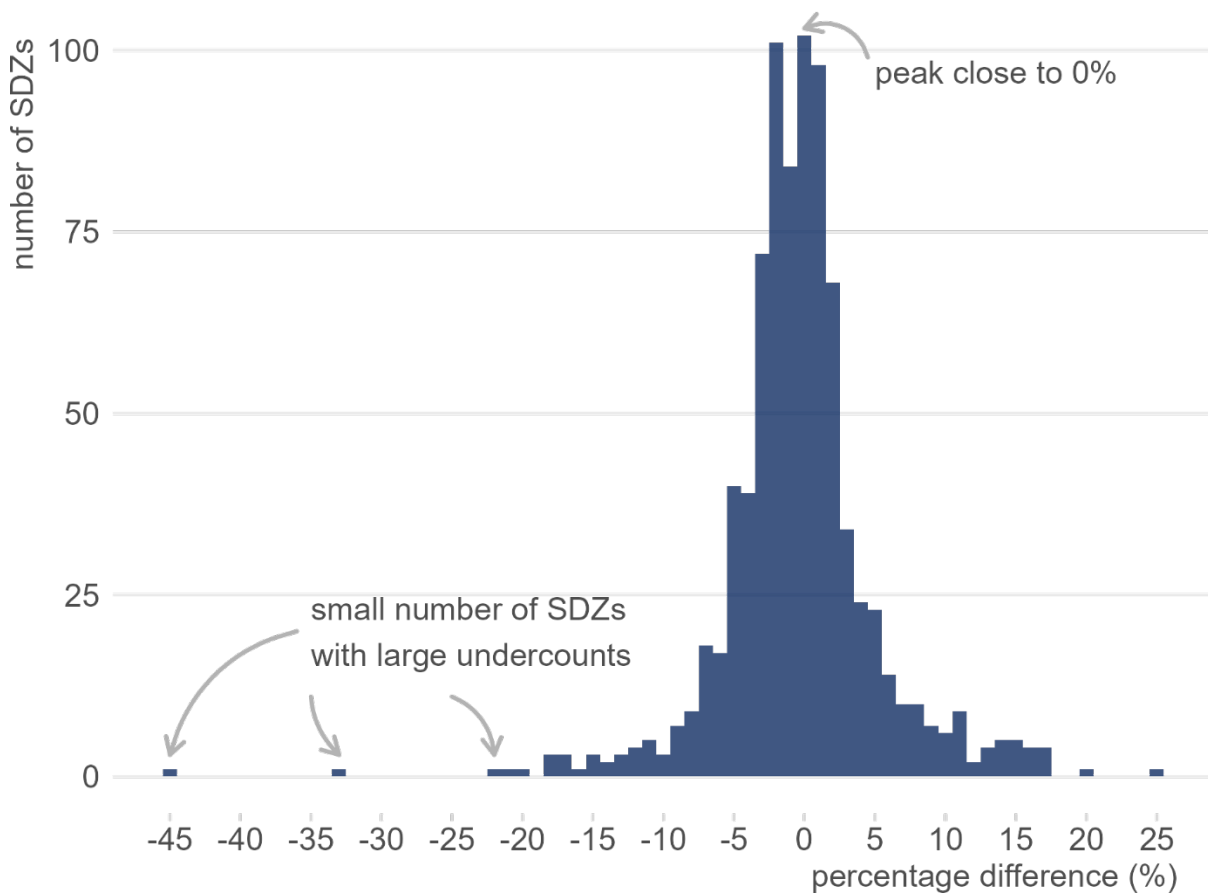
Threshold	Number of SDZs	Percentage of total SDZs
Within 2%	388	45.6%
Within 5%	657	77.3%
Within 10%	784	92.2%

[Download data](#) (xlsx, 54 KB)

The percentage differences between the SPD and census at the SDZ level were broadly normally distributed as can be seen in Figure 6. The mean percentage difference was -0.4% as SPD estimates tended to be slightly less than census populations. The standard deviation was 5.6 percentage points.

¹⁶ [Super Data Zones](#) (Census 2021) including information and digital boundaries.

Figure 6: Distribution of differences between SPD and census at SDZ level, 2021



[Download Chart](#) (xlsx, 140 KB)

These are larger percentage differences than were observed at the LGD level, but greater variability was expected as SDZs have much smaller populations than LGDs and the effects of population churn can have a greater impact on their populations. The smaller populations mean that relatively small differences in population resulted in larger percentage differences. SDZs that contained larger proportions of the types of people that were less likely to appear regularly on administrative datasets, such as younger working age adults, were more likely exhibit greater SPD undercounts.

Super Data Zone outliers

There were five SDZs with an absolute percentage difference between SPD and census of more than 20%. Four of these were SPD undercounts (Castle_Q, Botanic_A and Botanic_R in Belfast LGD; and Airport_F in Antrim and Newtownabbey LGD) and one was an overcount (Foyleside_F in Derry City and Strabane LGD). The four undercounted SDZs all appeared within the top ten SDZs with the highest proportion of usual residents with an address one

year ago outside of NI. At least half of the populations of Botanic_A (49.6%), Botanic_R (50.2%) and Castle_Q (57.2%) SDZs lived somewhere else one year ago according to census¹⁷. Each of these SDZs had a relatively large proportion of their usual resident population who were full time students above compulsory school age (49.9%, 39.2% and 49.5% respectively) in the census¹⁸. These factors indicate greater levels of population churn, and a more mobile population could have resulted in greater impacts on the SPD estimates due to the time lags in the administrative data, counting some people in the wrong location and missing other people altogether.

It is known from census that there were 13 SDZs with more than 10% of their usual residents living in a communal establishment (CE)^{13,19}. Almost all (12) of these SDZs had an SPD undercount and most (11) had an undercount of at least 5%. The SPD had a limitation when it comes to CE residents. The SPD may have undercounted because many of the residents did not appear recently in the administrative datasets used, or may appear in the administrative datasets at their home locations rather than CE locations. In contrast the census was more accurate because CE residents were counted via communal establishment returns or special population groups²⁰. See the [limitations section](#) for more information. Additional administrative data might help to reduce the impact of some of these limitations, but the CE issue might require a change in methodology or additional data collection.

The SDZ that the SPD overcounted the most, Foyleside_F, contains a university campus and a relatively high proportion of its population (10.1%) were full time students above compulsory school age in the census¹⁸. This was also true of the neighbouring SDZs Foyleside_D (18.2%) and Foyleside_G (10.9%). The population churn involving students could have caused addressing issues for the SPD, resulting in a particularly high overcount in the Foyleside_F SDZ.

¹⁷ Census 2021 Flexible Table Builder: [Address One Year Ago \(4 categories\) by SDZ](#)

¹⁸ Census 2021 Flexible Table Builder: [Compulsory School Age by Full Time Student by SDZ](#)

¹⁹ Census 2021 CT0099: [Communal establishment management and type](#) (xlsx, 369 KB)

²⁰ Census 2021 [Population Definitions for Northern Ireland](#): sections 2.3 (communal establishments) and 3.5 (special population groups).

4 Background Notes

4.1 Disclaimer

The SPD estimates do not replace the accredited official statistics: the Northern Ireland mid-year population estimates. The SPD uses a different methodology to the NISRA mid-year population estimates. Any use of these outputs should make clear to users the nature and purpose of the research and should include this disclaimer.

The SPD estimates are compared to the accredited official statistics for the Census 2021 in this research. The comprehensive nature of the census makes it crucial in a wide range of different settings. It is used as a baseline for the creation of annual population estimates and, due to the detailed picture that it provides of the population of Northern Ireland, provides a high quality benchmark for this research. This SPD has been created at April 2021 to enable comparison with the census.

The accredited official statistics for NI population estimates and census statistics are the mid-year population estimates and NI Census 2021 statistics respectively:

- [NISRA mid-year population estimates](#)
- [NISRA Census statistics](#)

NISRA intend to compare SPD to mid-year population estimates in future publications.

4.2 Statistical Research

NISRA consider these outputs to be statistical research rather than official statistics in development. This statistical research explores the creation of population estimates for NI using a range of administrative datasets. It describes how the SPD estimates were created, what they estimate the NI population to be in April 2021, and how they compare to the published census statistics (dated March 2021).

Statistical research is an appropriate label for the SPD at this time because it does not yet meet the requirements for official statistics in development²¹. NISRA would like to support

²¹ [Guidance on producing official statistics in development](#) from the Office for Statistics Regulation.

user understanding and the development of the SPD and inform discussion by publishing this statistical research, and the research to follow. It is also important that this research does not undermine the existing accredited official statistics for the NI population: the mid-year population estimates. NISRA continues to develop the methods and data used to create the SPD and will produce more evidence of the quality of the SPD. In early 2025, NISRA will publish a Quality Assurance of Administrative Data (QAAD) assessment for the SPD. The QAAD will provide reassurance and transparency on how the quality of the SPD outputs are affected by the data quality of the source datasets that are used to create it. It will give users a better understanding of the reliability and accuracy of the SPD. NISRA will produce further research to inform future developments in population and census statistics in 2025.

4.3 Census 2021

The Census 2021 statistics relate to census day, 21 March 2021. The census placed a legal obligation on every householder and every manager/resident of communal establishments to make a census return. Obligations were also placed on owners of vacant properties. Census returns could be made online, by a paper questionnaire, or via telephone.

- [NISRA Census 2021](#)

The analysis in this report considers different aggregations of the SPD by combinations of age, sex, Local Government District (LGD) and Super Data Zone (SDZ) and compares each aggregation to Census 2021. The [Census 2021 main statistics tables](#) used to make these comparisons are listed below alongside custom tables created using the Census 2021 [flexible table builder](#):

- Main Statistics MS-A01 - [Usual resident population](#)
- Main Statistics MS-A02 - [Age - five year age bands - all usual residents](#)
- Main Statistics MS-A05 - [Age - single year](#)
- Main Statistics MS-A07 - [Sex](#)
- Main Statistics MS-A08 - [Five year age bands and sex](#)
- Main Statistics MS-A09 - [Single year of age and sex](#)

- Flexible Table Builder: [Economic Activity by Health in General by LGD](#)
- Flexible Table Builder: [Compulsory School Age by Full Time Student by LGD](#)
- Flexible Table Builder: [Address One Year Ago \(6 categories\) by LGD](#)
- Flexible Table Builder: [Address One Year Ago \(4 categories\) by SDZ](#)
- Flexible Table Builder: [Compulsory School Age by Full Time Student by SDZ](#)
- Commissioned Table CT0099: [Communal establishment management and type](#)
- Commissioned Table CT0116: [Usual resident population by LGD and Urban Status](#)

4.4 Statistical Disclosure Control

Statistical Disclosure Control (SDC) refers to a range of methods that aim to protect individuals, households, businesses, and their attributes from being identified in published data and microdata. For Census 2021, NISRA have applied three strategies to ensure individuals are protected from identification while minimising the impact on the quality of results: targeted record swapping, cell key perturbation, and disclosure rules. For more information please see the Census 2021 guidance note: [statistical disclosure control methodology for 2021 Census](#).

The implications of the SDC in the census on this research, are that the small amounts of “noise” that were added to cell counts may result in small differences between SPD and census. This was not a significant limitation to the research as this analysis compares SPD and census and looks for population estimates within thresholds (usually 2% and 5%) rather than seeking counts that precisely match.

For the SPD, statistical disclosure control was applied at the table design stage to avoid the production of small population counts. The NI level data was presented by single year of age and sex; the LGD level data by five year age band and sex; and at SDZ level only population totals. This provided a balance between protecting the privacy of individuals while meeting the user need of providing a meaningful comparison.

This research compared the SPD and census population estimates, which came from two different methodologies. The differences were small for some combinations of age, sex and geography, including counts of 0, ± 1 and ± 2 . However, these small counts were not

considered to be disclosive. The statistical disclosure control methods applied and the uncertainty around the estimates, which were quantified for census⁵ but not for SPD, mean that small differences should not be taken to refer to particular individuals.

4.5 Mid-Year Population Estimates

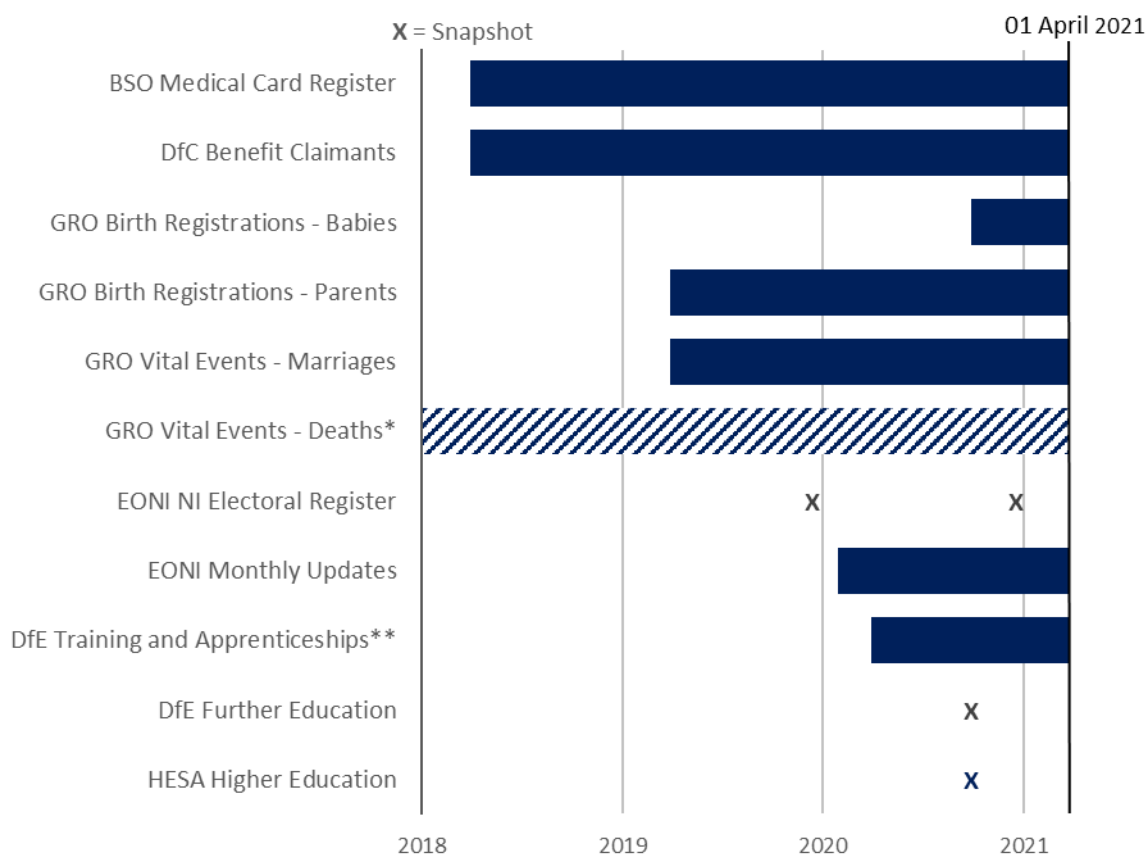
NISRA produce estimates of the population of NI as of 30 June each year by age and sex. These are published annually and approximately one year in arrears. Population estimates are based on the most recent census of population. Each year, the population is 'aged-on' by one year with the number of births in the year added and the number of deaths in the year subtracted. An adjustment is also made for migration. Subsequent mid-year population estimates then use the previous year's figures as the base.

- [NISRA mid-year population estimates](#)

4.6 Data Acquisition

NISRA have investigated a range of administrative data sources managed by different public bodies that could be used to estimate the NI population since before Census 2011. Discussions were held with each administrative dataset's owners and data sharing agreements were agreed to allow for the regular supply of these datasets. These legal agreements are [UK General Data Protection Regulation](#) compliant and are adhered to strictly by Census Office. The datasets used to create this SPD are shown in Figure 7.

Figure 7: Administrative data used to create the Statistical Population Dataset 2021



*GRO Deaths includes those deaths registered in NI since 1997.

**DfE Training and Apprenticeships includes participants on Training for Success, ApprenticeshipsNI and Skills for Life and Work programmes.

Note: See the [glossary](#) for the meanings of acronyms.

4.7 Data Processing

The administrative data are securely transferred from each data supplier to NISRA at regular intervals, usually monthly, quarterly or annually. Once a data extract is received, it is saved to a secure environment and quality assured. Each dataset extract follows a documented process which includes high-level checks, detailed variable-by-variable checks, and then processing of the variables. All data quality issues are logged, queried with the data supplier and resolved prior to the data being used.

4.8 Usually Resident Population

For the purposes of the census, a usual resident of the UK is anyone who, on 21 March 2021, is in the UK and has stayed or intends to stay in the UK for a period of 3 months or more, or has a permanent UK address and is outside the UK and intends to be outside the UK for less than 12 months²². Schoolchildren and students in full-time education with a term-time address that was different to their family home were counted as usually resident at their term-time address.

The SPD was created by combining the administrative datasets listed in Figure 7 at the individual record level. The SPD approximates the usually resident population by counting people that appear in the administrative datasets on the reference date and excluding people who were not present on the reference date.

The administrative datasets used to create the SPD were collected to support the functions of public bodies, and not for measuring the population. People who had recent interactions with the various administrative datasets in a period of time before the reference date, were considered to be present, unless the administrative dataset contained other information which indicated that the person was not currently resident. For example, deducted from medical card data due to emigration or death, or a higher education student enrolment with non-NI term-time postcode.

If a person had an interaction with one or more of the administrative datasets in an appropriate period of time before the reference date, that person was considered to 'have activity' or 'be active' in the administrative data and was included in the SPD. The period of time considered was called the 'activity window' and varied by dataset. How each administrative dataset was used varied and was determined by the inclusion and exclusion rules.

NISRA investigated different combinations and variations of these rules and the SPD estimates in this research paper show the results of the best estimate at the time of publication. The population estimates obtained depend on the methodology used; different

²² [Population Definitions for NI Census 2021](#) - Usual Resident definition on page 15

rules would have given different population estimates. These rules will be refined in future, as the available administrative datasets change and in response to the limitations identified in this research.

4.9 Creating the SPD

The administrative datasets were combined to create the SPD. The SPD is a list of people at the chosen reference date - for this research that date is 1 April 2021. The SPD contained one record per person and had variables for each person's demographic characteristics (name, date of birth, sex, postcode and unique property reference number (UPRN)). Inclusion and exclusion rules were applied so that only people believed to be resident in NI on the reference date were included. Inclusion rules were designed to include people who were present in NI at the chosen date. Exclusion rules were designed to remove people who were not present. These rules are described below.

The SPD was aggregated by age, sex and different geographic areas such as LGDs and SDZs. The resulting counts have been compared to equivalent Census 2021 population counts in this research paper. This section describes which administrative datasets and time periods were used and how they were combined.

4.9.1 Datasets and Inclusion Rules

The Business Services Organisation (BSO) provided information from the **medical card register** which is a comprehensive source of demographic information for NI. This administrative dataset included an activity variable to indicate whether a person had a recent interaction with primary healthcare services such as dental, ophthalmic or prescription services, or registered or changed address with their General Practitioner (GP). BSO's [Quality Assurance of Administrative Data Report](#) noted some weaknesses including a list discrepancy of approximately 5% which is relevant in the context of population counts such as this research. This was mitigated by combining the medical card register with other administrative data sources, and requiring a sign of recent activity.

Benefit claimant data was received from the Department for Communities (DfC) and was used in a similar way to the **medical card data**, to include people who had recent activity.

Births registrations data from the General Register Office (GRO) for NI provided data on births which was used to count babies born in NI in the previous six months. The aim was to count babies who may not yet have been registered with a GP on the medical card register. The individual record level data linkage reduced the chance of babies being double-counted.

Vital events data from GRO such as both partners from marriage registrations, and parents from births registrations, was used as indicators of recent activity. GRO death registrations were used to remove people who died before the reference date. This information may also have been captured in the medical card register data.

The Electoral Office for NI (EONI) provided the **NI Electoral Register** and monthly updates which were used to provide indicators of activity. If a person was added to the electoral register or contacted EONI to change their details, that interaction was considered a sign of activity. The dates used were determined by the availability of full electoral registers.

The Department for the Economy (DfE) provided **further education (FE) student enrolments and Training for Success (TfS), Apprenticeships NI and Skills for Life and Work programme participants** that were used to indicate recent activity. **Higher education student enrolments** from the Higher Education Statistics Agency (HESA) were used to indicate recent activity and were also used to move NI students to their term-time addresses. These datasets may have contained more recent activity data for young people who often are not interacting with the services covered by the other administrative datasets listed above.

4.9.2 Exclusion Rules

In some situations, people were excluded if they appeared in the medical card register, deaths registrations or higher education student enrolments data. People in the death registration data or on the medical card register with a date of death before 1 April 2021 were excluded. People who emigrated were flagged and those who had a deducted date after the latest activity date on other administrative data sources were excluded. Higher education student enrolments with a term-time postcode outside of NI were also excluded.

4.9.3 Address Rules

Once the inclusion and exclusion rules were applied, a series of address rules were applied to attempt to put each person into the correct address (UPRN). Due to the different purposes of the administrative datasets, the quality and timeliness of the address information contained within each administrative dataset varied. The address rules attempted to make the best use of the available administrative datasets, to count people at their correct addresses.

4.10 Limitations

It is important to note that this methodology reused administrative data for a statistical purpose that was not the originally intended use of the data. The administrative data sources used in this research are maintained by different public bodies for operational purposes, such as the provision of services. By combining the administrative data sources together, some of the limitations of the individual administrative data sources were overcome, and a combined dataset was created that was able to estimate the population in NI by age and sex with geographic breakdowns.

People may appear in the administrative datasets and the SPD but do not reside in NI on the reference date.

This could result in the SPD overcounting. The exclusion rules mean that only the medical card register, deaths registrations and higher education student enrolments administrative datasets could remove people from the SPD. People who had recent activity but have since emigrated from NI and did not inform their GP, may have continued to be counted in the SPD until they no longer met the activity related inclusion rules for the various administrative datasets that they appeared on. There were people registered for services in NI who were not residents but were entitled to access services, for example cross border workers. These people were removed from the SPD via their non-NI address. There may also have been people in some of the administrative data sources who were not residents but who were registered for services. The administrative data sources have eligibility rules and checks for this, so numbers were expected to be low, but unquantified.

People may reside in NI on the reference date but not appear on any of the available administrative datasets.

This could result in the SPD undercounting. While the administrative datasets were effective at capturing births and deaths, it was difficult to get timely administrative data to measure migration. People who moved to NI but were yet to have an interaction with one of the administrative datasets used to create the SPD will have been missed. These people may appear in the administrative datasets later once they access primary healthcare services, register to vote, receive a benefit²³, get married, become a parent, enrol in higher or further education or participate in the Training for Success, Apprenticeships NI or Skills for Life and Work programmes. This limitation could be mitigated by acquiring and incorporating more administrative datasets into the SPD.

People may reside in NI on the reference date, appear in the administrative datasets but not the SPD.

This could result in the SPD undercounting. More people could be included in the SPD by increasing the length of the activity windows in the inclusion rules. The activity windows varied by dataset. This can reduce the SPD undercount for particular groups and in particular areas, however it can also increase SPD overcount for other groups and areas. NISRA will continue to develop the SPD methodology to make best use of the available administrative data.

People may be correctly included in the SPD but at the incorrect location.

When the SPD counted people at the wrong address, the result could be an overcount in one location and an undercount in another location. A communal establishment (CE) is an establishment providing managed residential accommodation. Managed in this context means full-time or part-time supervision of the accommodation²⁴. A communal establishment resident is a person whose place of usual residence is in managed residential accommodation. This means any person who was living or expected to live in a CE for six

²³ Note that the benefits administrative data used to create SPD estimates includes most benefits but did not include Universal Credit.

²⁴ [NI Census 2021 definition of communal establishment](#) (HTML)

months or more²⁵. Types of CE include student halls of residence, hospitals, care homes, prisons and armed forces bases.

Depending on the type of interaction a person had with a CE, the person's address in the administrative dataset may have been their home address rather than the address of the CE on the reference date. This means that SPD and census estimates may have counted CE residents in different locations. Other CEs may contain people that are less likely to interact with local services, for example, being in a prison for an extended period may prevent a person from being active on the administrative datasets used in the SPD.

The SPD address rules attempted to use the most recent address information for each person across all of the available administrative datasets. However, if a person moved address but had not yet updated their address information on any of the administrative datasets, then the SPD would have assigned an inaccurate location for that person. Additionally, if a person had different addresses on different administrative datasets, the dataset with the most recent activity was typically used, but research resulted in some adjustments to this to prioritise benefits or medical card register for particular age groups. These methods can be changed in future to reflect changes in the data or new administrative datasets.

Data Quality and Availability

The administrative datasets used to create the SPD were high quality. NISRA will publish a Quality Assurance of Administrative Data (QAAD) assessment for the SPD in early 2025 to provide more information. However, a limitation of using administrative datasets to estimate population is that if data quality were to fall in one or more administrative datasets, it could diminish the quality of the SPD estimates. Furthermore, any delays to the availability of administrative datasets could cause delays in the production of the SPD.

The SPD combines the different administrative datasets using data matching²⁶. NISRA's data matching processes are well-researched and involve thorough quality assurance. It is a limitation of the SPD methodology that any issues relating to the data matching could result

²⁵ [NI Census 2021 definition of communal establishment resident](#) (HTML)

²⁶ [Data Matching Using NI Administrative Data: A Worked Example](#) (pdf, 650KB)

in quality issues in the SPD. For example, false positive matches could reduce estimated population counts by counting different people as the same person. False negatives are matches that should have been made but were not. Large numbers of false negatives could inflate SPD estimates by counting the same person twice.

4.11 Future Developments

NISRA are working to improve the SPD in the following ways.

Quality Assurance of Administrative Data

NISRA will publish a Quality Assurance of Administrative Data (QAAD) assessment for the SPD in early 2025. The QAAD will provide reassurance and transparency on how the quality of the SPD outputs are affected by the data quality of the source datasets that are used to create it. It will give users a better understanding of the reliability and accuracy of the SPD.

Additional administrative datasets could be acquired

Additional administrative data could address the specific quality issues identified in this research, for example the undercounting of young adults. The demographic details collected by the Department for Work and Pensions (DWP) and His Majesty's Revenue and Customs (HMRC) for the administration of the tax system may include the younger adults who were missing from this version of the SPD. Administrative data relating to migration could improve the quality of the SPD by including recent arrivals and excluding recent departures.

Improving the SPD using additional data would be an example of principle 1 (suitable data sources) of the quality pillar of the [Code of Practice for Statistics](#). NISRA are investigating the availability of such data.

Compare SPD to census at the individual level

Both the SPD estimates and Census 2021 published population counts were based on individual record level data. This research paper compared aggregated counts for different combinations of age, sex and geographic location. It would be possible to produce more detailed research comparing the SPD to Census 2021 if the analysis was at individual record level, rather than the aggregated data used in this research paper.

Compare SPD to mid-year population estimates

The mid-year population estimates are the accredited official statistics for NI. The analysis in this research paper can be repeated to compare SPD estimates to mid-year population estimates. This could be done for multiple years and for more recent years.

Coverage adjustment

The SPD does not currently adjust for undercoverage or overcoverage. NISRA could investigate coverage adjustment, for example using a coverage survey to inform dual-system estimation, as was done for Census 2021²⁷. A methodological advance like coverage adjustment would be an example of principle 2 (sound methods) of the quality pillar of the [Code of Practice for Statistics](#).

SPD time series

At the time of publication, the current SPD version could be produced for 2017 to 2022. The SPD methodology produces years that are independent from each other so errors in one year will not necessarily accumulate into the next year. However, there can be issues of coverage and quality in different administrative datasets that can change over time.

Produce SPD at other reference dates

This could be useful for supporting the accredited official statistics for population, the NI mid-year population estimates. These would benefit from an SPD with a reference date of 30 June rather than 1 April.

4.12 Link to Data used in this Research

This statistical research paper and the data used to create the commentary, tables, charts and map are available from the [research into using administrative data for population estimates](#) section of the NISRA website.

All of the data can be downloaded in a [single Excel spreadsheet](#) (xlsx, 308 KB) which contains several breakdowns by age, sex and geography; or as [separate spreadsheets](#) in the format used to create each of the separate tables, charts and maps.

²⁷ [Census 2021 data processing overview](#), Oct 2021

5 Enquiries and Suggestions

We welcome feedback from users on the content, format and relevance of this research paper. Please send feedback directly to population@nisra.gov.uk

Follow NISRA on [X](#), formerly known as Twitter, and [Facebook](#).

All media inquiries should be directed to the Department of Finance Press Office:

- Telephone: 028 9081 6724
- Email: dof.pressoffice@finance-ni.gov.uk

Further statistical information can be obtained by contacting:

- Telephone: 028 9025 5156
- Email: population@nisra.gov.uk
- Responsible Statistician: Gemma McGreevy

6 Glossary

BSO	Business Services Organisation (supports the Health and Social Care System in NI)
CE	Communal Establishment
CSO	Central Statistics Office (Ireland's national statistical institute)
DfC	Department for Communities (Northern Ireland)
DfE	Department for the Economy (Northern Ireland)
DoF	Department of Finance (Northern Ireland)
DWP	Department for Work and Pensions
EONI	Electoral Office of Northern Ireland
FE	Further Education
GDPR	General Data Protection Regulation (specifically UK-GDPR)
GP	General practitioner
GRO	General Register Office (for Northern Ireland)
HESA	Higher Education Statistics Agency
HMRC	His Majesty's Revenue and Customs
HTML	Hypertext Markup Language
KB	Kilobyte (file size)
LGD	Local Government District
MB	Megabyte (file size)
NI	Northern Ireland
NISRA	Northern Ireland Statistics and Research Agency
NRS	National Records of Scotland
ONS	Office for National Statistics
OSR	Office for Statistics Regulation
QAAD	Quality Assurance of Administrative Data
SDC	Statistical Disclosure Control
SDZ	Super Data Zone
SPD	Statistical Population Dataset
TFS	Training for Success
UK	United Kingdom of Great Britain and Northern Ireland
UPRN	Unique Property Reference Number