



Llywodraeth Cymru
Welsh Government

STATISTICS, DOCUMENT

Differences between estimates of Welsh language ability in Census 2021 and household surveys

Initial findings from research reviewing differences between estimates of Welsh language ability in Census 2021 and the Labour Force Survey. This publication is a product of a joint workplan between the Welsh Government and Office for National Statistics to improve our understanding of Welsh language statistics.

First published: 26 October 2023

Last updated: 18 December 2023

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Contents

[Introduction](#)

[Main points](#)

[Background](#)

[Agreement rates by Welsh language skill](#)

[Agreement rates by population characteristics](#)

[Agreement rates by Welsh-speaking frequency](#)

[Agreement rates by household type](#)

[Next steps](#)

[Methods and data quality](#)

[Notes on the use of statistical articles](#)

[Contact details](#)

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Introduction

In April 2023, the [Welsh Government and the Office for National Statistics \(ONS\) published a joint workplan](#). This workplan outlines the work we are planning to undertake during 2023-24 and beyond to improve our understanding of the main survey and administrative data sources used to produce statistics about the Welsh language. The workplan was accompanied by a [blog published by the Welsh Government's Chief Statistician](#).

This statistical article shares initial findings from the first of six projects outlined in the workplan. Project 1 involved analysing linked, de-identified Census 2021 and Labour Force Survey (LFS) data in the [Integrated Data Service \(IDS\)](#). The aim of the project was to better understand the characteristics of people who provided conflicting responses to the Welsh language question across the two sources.

This work project is an important part of assessing the current evidence base for Welsh language skills. [The Welsh language strategy, Cymraeg 2050](#), states that progress towards the Welsh Government target of a million Welsh speakers by 2050 will be monitored using census of population data. Findings from this project will help inform the Welsh Government's response to the National Statistician's recommendation following the [ONS' consultation on the future of migration and population statistics in England and Wales](#).

This work was supported by [Administrative Data Research \(ADR\) Wales](#) within Welsh Government and ONS colleagues from the Population Statistics team. Further information about the data-linking methodology and data quality can be found in the section on [Methods and data quality](#).

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Main points

- Around two in five people (39.9%) who recorded being able to speak Welsh in the LFS or Census 2021 recorded that they were unable to do so on the other source.
- More people reported that they were able to speak Welsh in the LFS and unable to do so in Census 2021 than vice versa.
- Among people who say they can speak Welsh in the LFS or Census 2021, the following groups tended to agree across both sources most often: people aged 65 years or older; people living in the north west; people born in Wales; and people with a Welsh national identity.
- Among people who say they can speak Welsh in the LFS or Census 2021, the following groups tended to disagree across both sources most often: people under 25 years old; people living in the south east and north east; people born elsewhere in the UK; and people without a Welsh national identity.
- Among people who agreed that they could speak Welsh on both sources, over two thirds (68.6%) reported that they spoke Welsh daily. Among people who said they could speak Welsh in the LFS but not in Census 2021, only around a quarter (24.8%) reported that they spoke Welsh daily.
- Proportionately fewer couple households agree on their Welsh-speaking ability between the two sources where none, or only one adult can speak Welsh compared with couple households that include two or more adults who can speak Welsh.

Background

Differences in the estimates of Welsh language ability between the census and household surveys such as the Annual Population Survey (APS), which uses data from multiple waves of the Labour Force Survey, are longstanding. Both the

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

ONS ('[Differences in estimates of Welsh Language Skills](#)' (The National Archives)) and the Welsh Government ('[Welsh language data from the Annual Population Survey: 2001 to 2018](#)') have explored possible reasons for some of these differences in the past.

While household surveys typically provide us with higher estimates of Welsh-speaking ability, this is the first time that the census has shown a decline in the number of Welsh speakers whilst the APS has shown an increase.

On 6 December 2022, the Welsh Government published a [statistical bulletin summarising the initial results from Census 2021 on the Welsh language skills of the population living in Wales](#). According to Census 2021, an estimated 538,300 usual residents aged three years or older in Wales were able to speak Welsh, or 17.8% of the population.

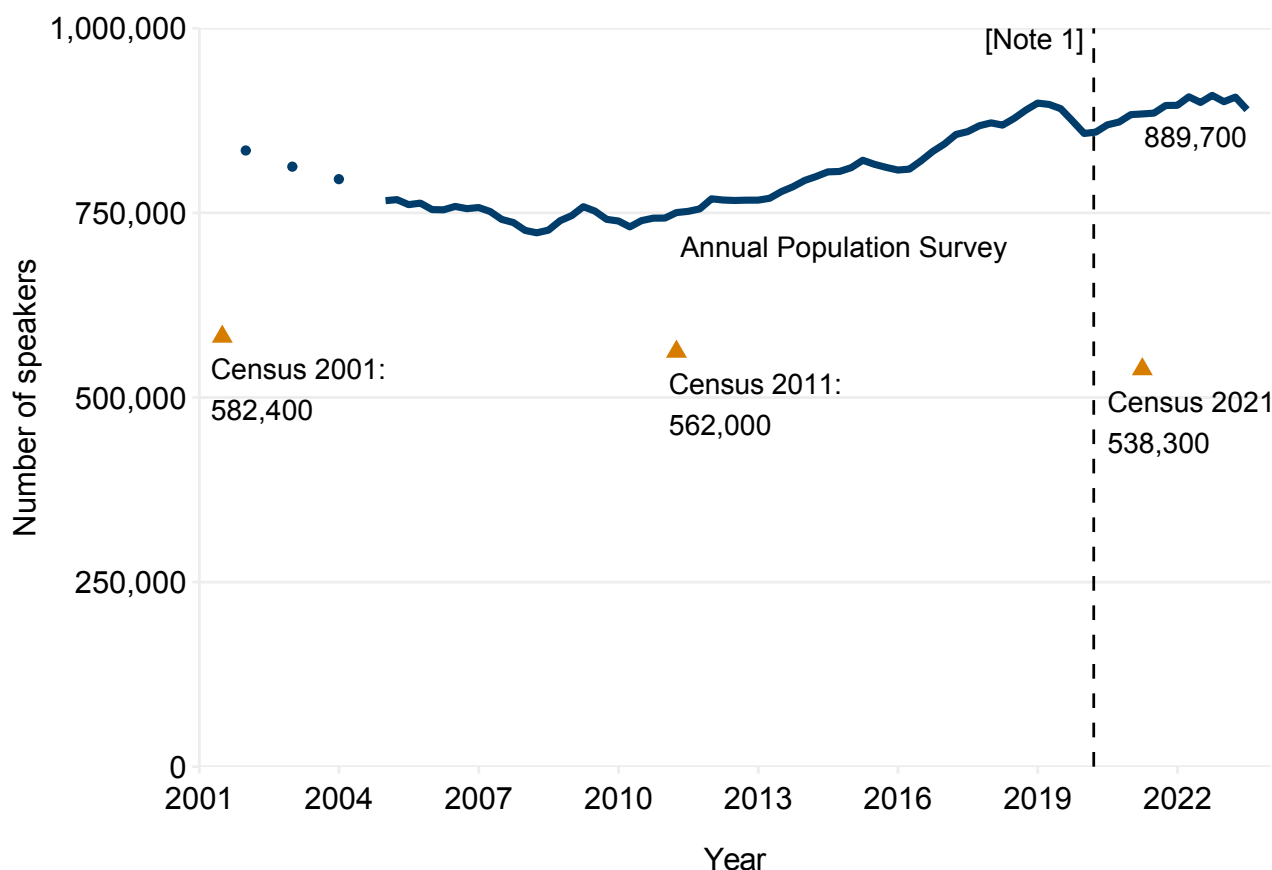
By comparison, at the time Census 2021 took place, [data from the Annual Population Survey](#) pointed to an estimated 884,000 Welsh speakers aged three years or older living in Wales (29.2% of the population), with a confidence interval of plus or minus 23,000 based on the results from the sample.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Figure 1: Number of people aged three years or older who are able to speak Welsh, 2001 to June 2023 [Note 1]



Description of Figure 1: This line chart shows that, having fallen between 2001 and 2007, there has since been an increase in the estimated number of Welsh speakers recorded by the APS. According to the APS, there were an estimated 889,700 Welsh speakers living in Wales in the year ending 30 June 2023. The number of Welsh speakers recorded in the 2001, 2011 and 2021 Census are plotted on the same chart, labelled 582,400, 562,000 and 538,300 respectively.

Source: Census of population (ONS); Annual Population Survey (ONS).

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Annual Population Survey: Welsh language data (StatsWales)

Census 2021: Welsh language data (StatsWales)

[Note 1] From mid-March 2020, the APS has been carried out by telephone only.

The APS estimates were higher than the Census 2021 estimates for each of the 22 local authorities in Wales. The largest differences can be seen in Blaenau Gwent, Newport, and Caerphilly, where the number of people able to speak Welsh recorded by Census 2021 was less than half of the APS estimate. The number of Welsh speakers recorded by Census 2021 in the Isle of Anglesey and Gwynedd were 18% and 19% lower than the APS estimates respectively. These were the local authorities with the smallest differences between the recorded estimates.

For nearly every local authority, the percentage point difference in the proportion of people able to speak Welsh between both sources was higher in 2021 than in 2011, suggesting that both sources are becoming increasingly different over time.

Similarly, the size of the difference between the two sources varies by age group. In both 2011 and 2021, the largest differences between the sources were seen for the youngest age groups (ages 3 to 15 and 16 to 24), with the highest estimate recorded by APS. These were also the age groups with the highest reported percentages of Welsh speakers according to both sources.

Once again, for each age group, the percentage point difference in the proportion able to speak Welsh between both sources was higher in 2021 than in 2011.

While some of the differences may be attributable to the **sampling and weighting methodology for the LFS** (something that will be explored further in future work), this article focuses on the extent to which the same individual's

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

responses differ between both sources.

Agreement rates by Welsh language skill

In this article, "headline agreement rates" refer to the proportion of individuals in the linked dataset who recorded the same response to a given question across both sources. Note that for some individuals, another person may have responded on their behalf, including about their Welsh language ability. In the LFS, another person always responds on behalf of children aged under 16 years.

The linked dataset used in this analysis contains 5,380 individual records which span five quarters of data from the LFS. Further information can be found in the section on [Methods and data quality](#).

When asked about their ability to speak Welsh, 87.3% of linked respondents gave a consistent answer in the LFS and Census 2021, while 12.8% gave conflicting responses.

Table 1: Linked respondents' recorded ability to speak Welsh in the Labour Force Survey (LFS) and Census 2021 [Note 1]

Ability to speak Welsh	Cannot speak Welsh (LFS)	Can speak Welsh (LFS)
Cannot speak Welsh (Census 2021)	68.1%	10.9%
Can speak Welsh (Census 2021)	1.9%	19.2%

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

[Note 1] The percentages recorded in this, and subsequent figures are calculated using rounded counts, and may not always total 100%.

The high percentage of consistent answers can be attributed to the fact that more than two-thirds of respondents recorded that they were unable to speak Welsh on both sources. In this article, the "conditional agreement rate" refers to the agreement rate when the individuals who record being unable to speak Welsh on both sources are excluded.

Among people who recorded an ability to speak Welsh on at least one source, the conditional agreement rate is 60.1%. Put another way, around two in five people who recorded being able to speak Welsh in the LFS or Census 2021 recorded that they were unable to do so on the other source.

Of the linked respondents who recorded that they were able to speak Welsh in the LFS:

- 63.8% also recorded that they were able to speak Welsh in Census 2021
- 36.2% recorded that they were not able to speak Welsh in Census 2021

Conversely, of the linked respondents who recorded that they were able to speak Welsh in Census 2021:

- 91.2% also recorded that they were able to speak Welsh in the LFS
- 8.8% recorded that they were not able to speak Welsh in the LFS

More people reported that they were able to speak Welsh in the LFS and unable to do so in the Census than vice versa. This result is expected given that the APS provides higher estimates of Welsh-speaking ability than the census.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

When we look at other Welsh language skills (ability to write, read or understand spoken Welsh), the headline agreement rate is somewhat higher for the least common skills, where a proportionately larger share of respondents agree that they do not have such a skill. When the respondents who record that they do not have a skill in either the LFS or Census 2021 are excluded, the conditional agreement rate is consistent across all four skills.

Among linked respondents in our dataset:

- 89.3% gave a consistent answer about their ability to write in Welsh (the conditional agreement rate was 59.8%)
- 87.8% gave a consistent answer about their ability to read Welsh (the conditional agreement rate was 59.1%)
- 84.8% gave a consistent answer about their ability to understand spoken Welsh (the conditional agreement rate was 59.7%)

We can also examine to what extent the combination of Welsh language skills recorded for linked respondents are consistent across the LFS and Census 2021.

Table 2: Percentage of Welsh language skills recorded by linked respondents in the LFS and Census 2021

Welsh language skills	No Welsh language skills (LFS)	Some Welsh language skills (LFS)	All Welsh language skills (LFS)
No Welsh language skills (Census 2021)	60.4%	6.6%	4.1%
Some Welsh language skills (Census 2021)	3.3%	4.4%	5.5%
All Welsh language skills (Census 2021)	0.7%	0.9%	14.2%

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

Table 2 shows a large degree of disagreement among people who recorded between one and three Welsh language skills – some, but not all – in the LFS. Among this group, 36.7% recorded having some but not all Welsh language skills in Census 2021, while the majority (55.5%) recorded having no skills at all.

Meanwhile, of the people recorded as being able to speak, read, write, and understand spoken Welsh in the LFS, 59.8% also recorded having all four skills in Census 2021.

Table 3: Most common disagreements about Welsh language skills recorded by linked respondents in the LFS and Census 2021

Census 2021	LFS 2021	Percentage of all disagreements
No Welsh language skills	Understand, speak, read and write Welsh	17.5%
No Welsh language skills	Understand spoken Welsh only	10.7%
No Welsh language skills	Understand and speak Welsh only	9.1%
Understand spoken Welsh only	No Welsh language skills	7.5%
Understand spoken Welsh only	Understand, speak, read and write Welsh	6.0%
All other skill combinations	All other skill combinations	49.2%

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

A more detailed examination of the combination of skills recorded by linked respondents shows that 76.6% noted the same four-skill combination in the LFS and Census 2021. Of the remaining responses, the most common disagreement involved recording no Welsh language skills in Census 2021 and all Welsh

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

language skills in the LFS (these account for 17.5% of all disagreements). The second most common disagreement involved recording no Welsh language skills in Census 2021 and the ability to understand spoken Welsh only in the LFS.

Together, the five most common disagreements listed in Table 3 account for over half of all observed disagreements about Welsh language skills in the linked data.

Agreement rates by population characteristics

Using responses provided to other questions in Census 2021, we can measure the agreement rates for Welsh language skills by a range of demographic and household characteristics. Since increasing the number of Welsh speakers is a key aim of the [Welsh Government's Welsh language strategy, Cymraeg 2050: A million Welsh speakers](#), the analysis in this section focuses on how respondents record their Welsh speaking ability specifically.

Agreement rates by age group

The vast majority of respondents in each age group agree about assessments of their Welsh speaking ability in both sources. However, the proportion who disagree varies considerably by age.

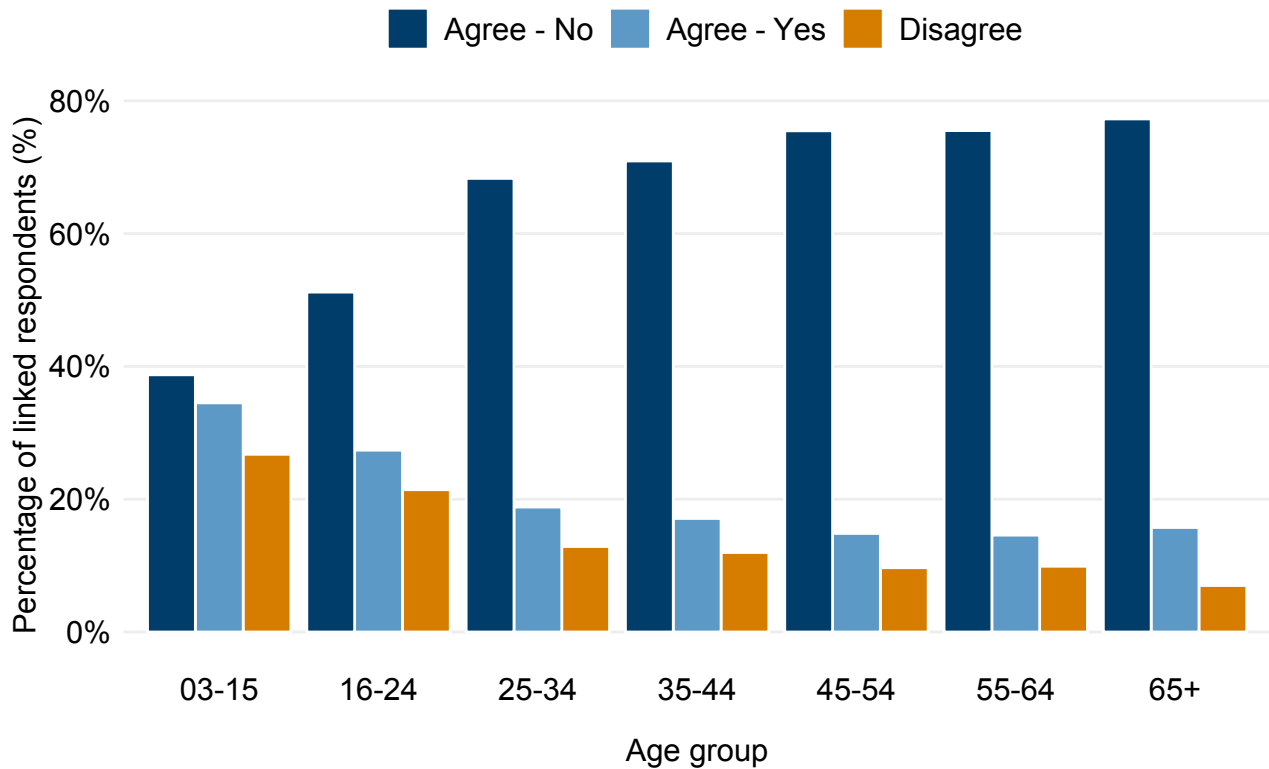
The proportion of disagreeing responses for children aged 3 to 15 years old was nearly four times as large as the equivalent proportion among people aged 65 years or older. It should be noted that in the LFS, another person always responds on behalf of children aged under 16 years.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

Figure 2: Percentage of linked respondents who agree / disagree about their Welsh-speaking ability in the LFS and Census 2021, by age group



Description of Figure 2: This grouped column chart shows the percentage of linked respondents in each age cohort who gave responses that disagreed about their Welsh speaking ability, agreed that they can speak Welsh, and agreed that they cannot speak Welsh in the LFS and Census 2021. A proportionately larger share of responses recorded for the younger age cohorts disagree, or agree that they can speak Welsh on both sources, compared with responses for their older counterparts.

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

Looking at responses recorded for children aged 3 to 15 years old:

- 26.8% disagree about their Welsh speaking ability in the LFS and Census 2021
- 34.5% agree that they can speak Welsh on both sources
- 38.7% agree that they are unable to speak Welsh on both sources

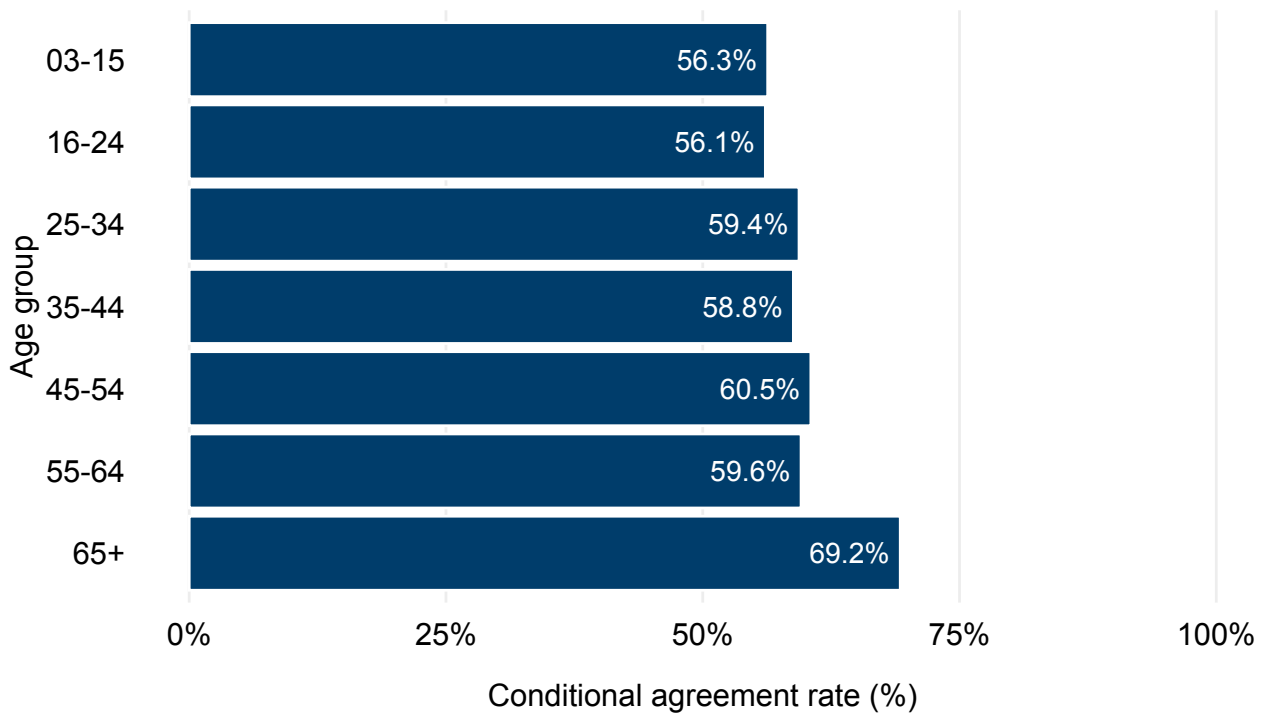
Among people aged 65 years or older:

- 7.0% disagree about their Welsh speaking ability in the LFS and Census 2021
- 15.7% agree that they can speak Welsh on both sources
- 77.3% agree that they are unable to speak Welsh on both sources

We know from census and APS data that the percentage of the population able to speak Welsh is highest among children and younger adults. Disagreement implies that the respondent must have been recorded as being able to speak Welsh on one source. The high prevalence of "disagreeing" and "agree - yes" respondents among the youngest two age cohorts might therefore simply reflect the age profile of Welsh speakers, **which is younger than that of the general population.**

To control for differences in the prevalence of Welsh speakers across age groups, we also measure the conditional agreement rate for those who say they can speak Welsh on at least one source.

Figure 3: Conditional agreement rate among people who record being able to speak Welsh on at least one source (LFS / Census 2021), by age group



Description of Figure 3: This bar chart shows that, of the people who say they can speak Welsh on at least one source, a proportionately larger share of people in the oldest age cohort were recorded as being able to speak Welsh on both sources than people in the youngest age cohorts.

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

Among people aged 65 years or older who say they can speak Welsh in the LFS or Census 2021, 69.2% also recorded they were able to do so on the other source. This conditional agreement rate is higher than for any other age group.

By contrast, children aged 3 to 15 years old and young people aged 16 to 24 years old have the lowest conditional agreement rates at 56.3% and 56.1% respectively.

Agreement rates by sex

Although a slightly greater proportion of women recorded their Welsh-speaking ability differently in both sources than men, the conditional agreement rates for men and women, which take into account that women make up a larger fraction of Welsh speakers, are broadly similar.

Of the linked respondents recorded as being able to speak Welsh on at least one source:

- 60.9% of women recorded being able to speak Welsh on both the LFS and Census 2021
- 59.7% of men recorded being able to speak Welsh on both the LFS and Census 2021

Agreement rates by region

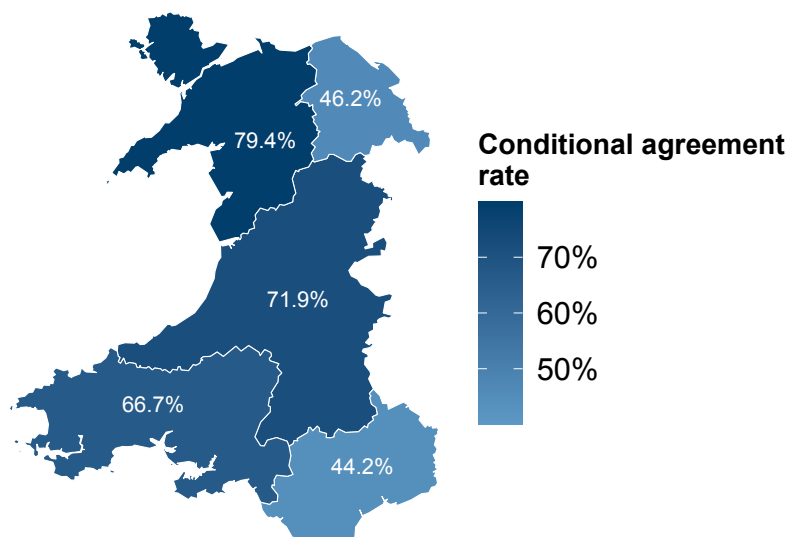
There are substantial differences in the conditional agreement rates by region. The five regions of Wales used in this analysis are consistent with those used in the National Survey for Wales's analyses of the Welsh language and the Welsh Language Use Survey 2019-20. **Further information about the mapping of local authorities to regions can be found in the most recent statistical bulletin on the Welsh Language Use Survey.**

In south east and north east Wales, the conditional agreement rates are 44.2% and 46.2% respectively. Put another way, of the people who said that they could speak Welsh in the LFS or Census 2021, over half recorded that they were

unable to speak Welsh on the other source. In both regions, there are more people who disagree than agree that they can speak Welsh on both sources.

In contrast, the highest conditional agreement rates are found in the north west (79.4%), followed by mid Wales (71.9%), and the south west (66.7%).

Figure 4: Conditional agreement rate among people who record being able to speak Welsh on at least one source (LFS / Census 2021), by region



Description of Figure 4: This map shows that the conditional agreement rate for people who say they can speak Welsh on at least one source is highest in the north west (79.4%) and lowest in the south east (44.2%).

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

The map broadly mirrors the geographic distribution of Welsh speakers, with the conditional agreement rate highest in areas where the largest proportion of the

population are able to speak Welsh. Data from the [Welsh Language Use Survey](#) also shows that areas with a higher proportion of Welsh speakers are strongly correlated with those areas where the language is most frequently used in [social settings](#) and in [the workplace](#).

Agreement rates by country of birth

Among people born in Wales, 63.4% of those who say they can speak Welsh in the LFS or Census 2021, say they can speak Welsh on both sources. For people born elsewhere in the UK, this conditional agreement rate is 43.8%.

It should be noted that people who say they were born elsewhere in the UK are disproportionately located in the local authorities which border England, including Powys and Flintshire. This group may include individuals who were born in hospitals in England but nevertheless grew up in Wales.

Agreement rates by national identity

Census 2021 also ask respondents about their national identity. Respondents can select one or several identities. In this article, we group respondents into two categories – people who select a Welsh national identity (which includes those who select Welsh and another identity), and people who do not select a Welsh identity.

Among people with a Welsh national identity, 64.9% of people who say they can speak Welsh in the LFS or Census 2021 say they can speak Welsh on both. For people without a Welsh national identity, this conditional agreement rate is 39.1%.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

Agreement rates by Welsh-speaking frequency

The LFS also asks respondents how often they speak Welsh. We used responses to this question to gauge whether the frequency of speaking Welsh is related to whether respondents agree about their Welsh speaking skills in the LFS and Census 2021.

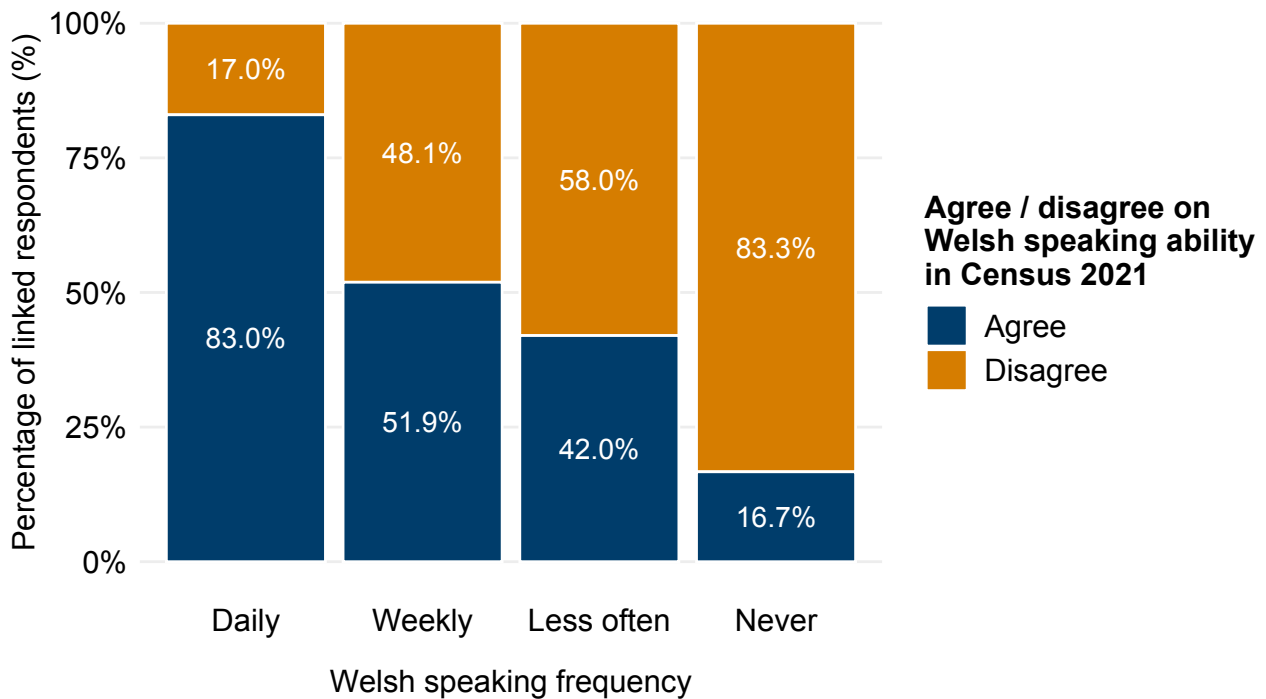
This question is not asked in census and is only prompted to those who say they can speak Welsh in the LFS. Therefore, when we refer to people who disagree in this section, we exclusively refer to people who say they can speak Welsh in the LFS but not in Census 2021.

Among people who agreed that they could speak Welsh on both sources, over two thirds (68.6%) reported that they spoke Welsh daily. Among people who said they could speak Welsh in the LFS but not in the census, only around a quarter (24.8%) reported that they spoke Welsh daily.

Put another way, of the linked respondents who say they can speak Welsh in the LFS, and do so daily, more than four in five reported that they can also speak Welsh in Census 2021.

Conversely, of those who say they can speak Welsh in the LFS, but never do so, more than four in five reported that they cannot speak Welsh in Census 2021.

Figure 5: Percentage of linked respondents who say they can speak Welsh in LFS who agree / disagree about their Welsh speaking assessment in Census 2021, by Welsh speaking frequency



Description of Figure 5: This stacked bar chart shows that, among people who say that they can speak Welsh in the LFS, most people who speak the language daily or weekly also report being able to speak Welsh in Census 2021. Conversely, most people who say they speak the language less often or never in the LFS report that they cannot speak Welsh in Census 2021.

Source: Census 2021 (ONS) and Labour Force Survey (ONS) linked dataset, April 2020 to July 2021.

Agreement rates by household type

Since all members of the household are invited to participate in the LFS and census, we also explore agreement rates at a household level.

In this section, a household is said to agree on its assessment of Welsh-speaking ability if all members of the household respond consistently in the LFS and Census 2021. This does not require all household members to give the same response, only that the response for any given member of the household is the same in the LFS and Census 2021.

Looking at the household agreement rate by broad household composition:

- 61.1%(r) of households with children agree on their Welsh-speaking ability in the LFS and Census 2021
- 83.8%(r) of households without children agree on their Welsh-speaking ability in the LFS and Census 2021

(r) These data items were revised on 18 December 2023. This rectifies a previous error whereby some input codes relating to the census household composition variables were incorrectly dropped, rather than mapped to the correct output codes.

Proportionately fewer households agree on their Welsh-speaking ability where none, or only one adult can speak Welsh (according to Census 2021) compared with households that include two adults who can speak Welsh. This possibly suggests that, where couples can speak Welsh to one another at home, household members have a clearer and more consistent understanding of their ability to speak Welsh.

Among couple households (either married, civil partnered or cohabiting):

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

- 71.7% of households where no adults can speak Welsh agree on their Welsh-speaking ability in the LFS and Census 2021
- 63.9% of households where one adult can speak Welsh agree on their Welsh-speaking ability in the LFS and Census 2021
- 82.8%(r) of households where two or more adults can speak Welsh agree on their Welsh-speaking ability in the LFS and Census 2021

r) This data item was revised on 18 December 2023. The previously published agreement rate had been derived by aggregating rounded counts for two output codes rather than aggregating the counts before rounding.

Next steps

While these initial findings provide some important insight into current data collection methods for Welsh language statistics, there is still more research that could be conducted to fully understand these differences.

Applying demographic weights to account for the differences between LFS respondents and the wider census population would enable a more robust interpretation of this analysis. This would also enable more direct comparison with existing [Census Quality Survey \(CQS\) findings on the ONS website](#).

To provide more context for this analysis, we could study census-LFS agreement rates for other variables. Again, this could be cross-referenced with CQS findings to identify whether observations are specific to the Welsh language questions, or more general to the nature of the LFS.

More complex analysis could also be undertaken to assess the significance of factors which lead to inconsistent reporting of Welsh language skills. This might be carried out through statistical modelling using multinomial regression.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

The analysis in this article found that a large fraction of disagreements involved data for people aged under 16 years. All LFS responses (and most Census 2021 responses) for this age group are provided by an adult in the household on behalf of the child in question. The proxy nature of these responses may therefore contribute to their lower agreement rates. We could analyse and compare the relationship between proxy response and agreement rates across Census 2021 and LFS for other age groups to understand this further.

Another area for research is Welsh language skills responses collected by the Transformed Labour Market Survey (TLFS). The **ONS currently plans to transition to the TLFS from the LFS to produce regular labour market and productivity outputs** by March 2024. TLFS, like the LFS, is a voluntary sample survey but, unlike the LFS, is conducted through online self-completion questionnaires.

The findings of this project will feed into other projects outlined in **the workplan**. This includes project 4, which explores how survey mode and design effects might impact the collection of information about Welsh language ability, and project 2, which will explore differences in reporting of Welsh language ability between **the Pupil-Level Annual School Census (PLASC)** and other data sources.

Regular updates will be provided on the progress of these and other projects as part of our quarterly publication on **Welsh language data from the Annual Population Survey**. The most recent update was published on 5 October 2023.

8.1 Consultation on the future of migration and population statistics in England and Wales

The ONS has been conducting a consultation on ambitious proposals for the future of population and migration statistics in Wales and England. These **proposals aim to put administrative data sources at the core of the system**

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

for producing population statistics, complemented by survey data and a wider range of other sources (ONS). This would provide more timely, high-quality statistics and may replace the current reliance on, and need for, a census every ten years.

The census in Wales has provided statistical estimates on Welsh speakers for more than a century. The current vision set out by the ONS means there is a need to understand the impact that moving away from a 'traditional census' may have on Welsh language statistics, and what benefits and opportunities for more timely, high-quality Welsh language data may exist within the proposed statistical system.

The ONS and the Welsh Government recognise the importance of Welsh language statistics to statistical users in Wales, and therefore want to understand the differences, strengths, and limitations of existing sources to help shape how these data will be produced in the future. This work project is an important part of assessing the current evidence base for Welsh language skills and will in part inform the Welsh Government's response to the National Statistician's recommendation following the consultation.

This research is particularly important as it is acknowledged that, at present, no comprehensive administrative data source exists that provides reliable data on Welsh speakers. The ONS recognises that alternative means of collection may be needed as part of the transformed system, for example, through government departments' administrative data collection or through new or existing surveys.

Additionally, it is acknowledged that at present there are no high-quality data on Welsh speakers outside of Wales that might provide an accurate picture on Welsh speakers in England or the rest of the UK. At present therefore, there are no robust sources that provide insight into Welsh speakers who have moved from Wales either temporarily (for example, to study) or more permanently. There is a recognition of this gap in the data and the need to consider how this might be addressed.

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

Methods and data quality

Integrated Data Service (IDS)

Research and analysis for this work project has been conducted in collaboration between the ONS and Welsh Government using the Integrated Data Service (IDS). The IDS is a cross-government project delivered by the ONS.

Building on the success of the Secure Research Service (SRS), the IDS is a multi-cloud platform that brings together ready-to-use data to enable faster and wider collaborative analysis for the public good. The IDS is a safe and secure data provisioning platform, a Trusted Research Environment compliant with **the five safes of secure data (IDS)**, and formally accredited as data provider under the **Digital Economy Act 2017 (DEA) (GOV.UK)**.

This project has been conducted in the IDS as an 'early adopter' project, and in part showcases the potential and capability of the service to enable cross-cutting collaborative research across government departments. This project has been approved by the **UKSA Research Accreditation Panel**. The analysis has been produced by ONS and Welsh Government researchers with remote safe researcher accreditation.

The **ADRW Programme of Work 2022 to 2026** outlines the ten thematic areas that the ADR Wales team will focus their research on to help government address the most pressing issues facing society. ADR Wales is part of ADR UK and funded by the Economic and Social Research Council (part of UK Research and Innovation).

Data linkage

The basis for the dataset used for this project is the Census Non-Response Link

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

Study 2021 (CNRLS). The CNRLS matched households and the people within them who completed an ONS survey around the time of Census 2021 with their corresponding census response. The methods used for matching persons and households were similar to the approach used for [matching Census 2021 to the Census Coverage Survey \(CCS\) \(ONS\)](#).

The dataset used for this analysis comprises the households and individuals matched through CNRLS 2021 who:

- Were in a cohort of the Labour Force Survey (LFS) actively sampled between January 2021 and June 2021;
- Responded to at least one wave of LFS between April 2020 and July 2021; and
- Responded to Census 2021

The LFS match rate to CNRLS, after automated and clerical matching, was very high. 99.5% of households, and 92.9% of individuals who responded to the LFS in England and Wales were matched to the CNRLS. The precision of matching was also very high; the false positive rate overall for CNRLS 2021 was 0.058% for address-level matching, and 0.044% for person-level matching. The false negative rate for address-level matching was 0.090%, and 0.072% for person-level matching.

Data quality

As the analysis of the linked dataset is based on unweighted response data, it is not representative of the census population for Wales. The LFS has experienced a decline in response rates during and after the pandemic, due to the mode change from face-to-face to largely telephone interviewing in the first wave. This has increased the risk of non-response bias in the survey. [The ONS has conducted analysis exploring the nature of these biases](#), and has made adjustments to the LFS weighting strategy to include [tenure weights \(ONS\)](#) and

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

use [HMRC Real Time Information \(ONS\)](#).

The linked dataset combines five quarters of data from the LFS to boost the sample size. Since survey respondents are invited to participate in five waves, each respondent may have provided up to five recorded responses to the questions on Welsh language skills on the dataset. Therefore, we use the response given in Q1 2021, if available. This is the period that contains Census Day, which was 21 March 2021. Otherwise, we use data from Q2 2021, or if this is not available either, we take the response given in the quarter closest to March 2021 chronologically. In our linked dataset, 58.6% of records have a response to the Welsh language skills questions in Q1 2021, while a further 35.3% have responses from the quarter immediately preceding or following Q1 2021.

The [ONS has previously published agreement rates for variables from Census 2021](#) by linking census records with responses to the [Census Quality Survey \(CQS\) \(ONS\)](#). Unlike our linked dataset, these agreement rates are calculated using demographic weighting to be representative of the census population.

The CQS agreement rate exceeds 98% for some straightforward variables (for example, age, sex, and country of birth). However, it is often lower, especially for questions that are subjective or changeable in nature. This includes general health (66.6%) and national identity (59.2%), the latter of which allows combinations of multiple options to be selected.

The published, weighted CQS agreement rate for overall combinations of all four Welsh language skills is 76.6%. The unweighted headline agreement rate between Census 2021 and the LFS at this level also happens to be 76.6%.

Where sample sizes permitted, we reviewed CQS data for Welsh language skills and found that results were broadly similar to the LFS-based comparison in this article. For example, both studies found that the vast majority of disagreements

involve respondents claiming a greater number of skills on the comparator survey than census. Similarly, both studies also found that over 90% of respondents who said that they could speak Welsh in the census agreed on the comparator survey.

These projects therefore corroborate and reinforce one another's findings on this topic.

Notes on the use of statistical articles

Statistical articles generally relate to one-off analyses for which there are no updates planned, at least in the short-term, and serve to make such analyses available to a wider audience than might otherwise be the case. They are mainly used to publish analyses that are exploratory in some way, for example:

- introducing a new experimental series of data
- a partial analysis of an issue which provides a useful starting point for further research but that nevertheless is a useful analysis in its own right
- drawing attention to research undertaken by other organisations, either commissioned by the Welsh Government or otherwise, where it is useful to highlight the conclusions, or to build further upon the research
- an analysis where the results may not be of as high quality as those in our routine statistical releases and bulletins, but where meaningful conclusions can still be drawn from the results

Where quality is an issue, this may arise in one or more of the following ways:

- being unable to accurately specify the timeframe used (as can be the case when using an administrative source)
- the quality of the data source or data used
- other specified reasons

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).

However, the level of quality will be such that it does not significantly impact upon the conclusions. For example, the exact timeframe may not be central to the conclusions that can be drawn, or it is the order of magnitude of the results, rather than the exact results, that are of interest to the audience.

The analysis presented does not constitute a National Statistic but may be based on National Statistics outputs and will nevertheless have been subject to careful consideration and detailed checking before publication. An assessment of the strengths and weaknesses in the analysis will be included in the article, for example comparisons with other sources, along with guidance on how the analysis might be used, and a description of the methodology applied.

Articles are subject to the release practices as defined by the [release practices protocol](#), and so, for example, are published on a pre announced date in the same way as other statistical outputs.

Contact details

Statisticians: Cian Siôn (Welsh Government) and Rob Doherty (Office for National Statistics)

Email: welshlanguagedata@gov.wales

Media: 0300 025 8099



This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys-html> for the latest version.

Get [information on copyright](#).

This document may not be fully accessible.

For more information refer to our [accessibility statement](#).

This document was downloaded from GOV.WALES and may not be the latest version.

Go to <https://www.gov.wales/differences-between-estimates-welsh-language-ability-census-2021-and-household-surveys.html> for the latest version.

Get [information on copyright](#).