CONSTITUTION, EUROPE, EXTERNAL AFFAIRS AND CULTURE COMMITTEE

19th Meeting, 2022, Session 6

8 September 2022

Inquiry into Scotland's Census

1. The Committee is taking evidence on Scotland's Census from two members – Professor Sir Ian Diamond and Professor David Martin – of the International Steering Group of global census experts established by the Registrar General.

2. There is a briefing from the International Steering Group at **Annexe A**.

3. A brief summary of the written submissions received in response to the Committee's call for views is available at **Annexe B**.

4. A paper prepared by SPICe on the approaches being taken to check and improve data quality can be found at **Annexe C**.

CEEAC Committee Clerks
June 2022

Annexe A

Constitution, Europe, External Affairs and Culture Committee: Inquiry into Scotland's Census

Response from the International Steering Group (https://www.scotlandscensus.gov.uk/about/2022-census/international-steering-group/)

1. The International Steering Group was formed in May 2022 by the Registrar General following a discussion with the UK's National Statistician and with the agreement of Scottish Ministers.

2. The aim was to convene a steering panel with a number of experts with backgrounds in censuses and the use of administrative data to consider the position with Scotland's Census.

3. The group was asked to help steer the work of National Records of Scotland as the Census moved out of collection phase and into the estimation and processing work required to take census records and convert them into high quality census outputs.

4. Specifically, the group's role in Scotland's Census 2022 is to:
    a. provide assurance that the census programme was ready to move on from the collect phase
    b. provide direction and support to National Records of Scotland as they implement their statistical design after the collection phase
    c. propose amendments or changes for NRS to consider, including accessing additional administrative data

5. This response has been prepared by the group based on its work since its first meeting on 27th May 2022. At that stage census response stood at 87.5%, and an early action of the group was to affirm that NRS should move from extended main phase enumeration to Census Coverage Survey (CCS) fieldwork. This was completed at the beginning of June with guidance from the International Steering Group that NRS should proceed to the next phase of the census process.

6. The main group initially met weekly with NRS, and then as work on the CCS progressed, moved to a fortnightly pattern with a weekly sub-group to discuss the CCS. The main group has focused attention on the methodology and next steps involved in the production of Scotland's 2021 Census estimates, with a

particular emphasis on the use of administrative data in quality assurance and estimation.

7. Relating to the Inquiry's exploration of broad census quality, the group's remit does not extend to investigation into the reasons driving response rates or the performance of specific census questions, except in so far as they can help us to fulfill our primary role of population coverage.

8. Minutes from the group's meetings will be made available on the NRS website.

Background

9. Modern censuses are no longer a simple count of the population, but a part of the wider system of population statistics. A combination of data sources are brought together **to provide statistical estimates** – that is, the census outputs are statistical estimates which have uncertainty, much like a sample survey such as the Labour Force Survey. At this point in time, NRS are only part-way through the process of producing the census outputs. While the main enumeration phases are complete, it is the entire process combining across multiple data sources that drive the quality of the final outputs.

10. The quality (or accuracy) of the census estimates is not purely a function of the headline census response rate but depends on the other data sources and the methodology used to combine them to provide the estimates. For context, the 2011 Census outputs include around 6 per cent of statistically imputed records which were estimated to have been missed through this type of methodology.

11. International approaches to census taking, including in all nations of the UK, recognise the importance of three pillars in the production of final census outputs.

1. An enumeration targeting coverage of, in the case of Scotland, all usual residents on Census Night at the location of their usual residence.
2. An independent CCS in the weeks following the census enumeration, targeting coverage of all usual residents on Census Night at the location of their usual residence for a sample of areas.
3. Use of administrative data in the quality assurance of the final census outputs.

12. Pillar 1) remains the foundation of the census outputs, and if 100% response was achieved nothing else would be required. Recognising that 100% response is in reality unachievable, Pillar 2) provides the data to estimate and adjust the enumerated counts for the (under-) coverage. This was done in Scotland in 2001 and 2011 Censuses. Pillar 3) then plays the role of providing comparison data for quality assurance, and in some cases information to correct for missing items of data within the responses provided under Pillar 1).

13. These pillars come together in a statistical estimation methodology which aims to firstly estimate each council area of Scotland by age-sex population size. These are the headline census population estimates (as noted above they are not counts). The combination of the Census, the CCS and the administrative data drive the quality of those key estimates. Specifically addressing the Inquiry's objectives, the estimation methodology is designed to draw on all three Pillars to fill the gaps in data collection.

14. The census database is adjusted, by imputing whole household and person records into each council area so that the number of records equals those estimates. The census output tables are drawn from that adjusted database, providing estimates (not counts) in all census outputs. An overview of the plans for Scotland's Census can be seen at https://www.scotlandscensus.gov.uk/about/2022-census/statistical-methodology/.

15. Both the headline census population estimates and the adjusted database are key components in the wider system of population statistics. The census population estimates are the primary source of information for bench-marking the ongoing series of population estimates, which are themselves bench-marks for a wide range of survey outputs, while the adjusted database provides inputs into local planning decisions and provides the basis to estimate more detailed population structures throughout the inter-censal period.

Current Estimates of Household Coverage for the Census Enumeration

16. Management information at the end of the census enumeration indicates an overall response of around 90%, with the lowest coverage for a council area at around 85%. This compares unfavourably with 2021 response in England and Wales of 97%, but is well within the bounds of recent international experience with New Zealand achieving 83.3% response in 2018 prior to enhancing their data with the inclusion of administrative data returns.

17. While the overall response for Scotland is lower than the original target of 94%, it is important to recognise that 30 of the 32 council areas did make the original minimum target of 85%, the final two remaining slightly below the 85% target. However, within those response rates there remains variation at local levels.

18. The variation needs to be estimated and adjusted to ensure the final census outputs deliver both the national picture for Scotland and the local-level picture essential for planning and service delivery.

Role of the Census Coverage Survey (CCS)

19. In both 2001 and 2011, the CCS was the key source of independent data to both estimate the response to the census and adjust the final census outputs as summarised above. Administrative data were used only for additional quality assurance comparisons.

20. Imputing missed households and individuals into the census ensures the local-level outputs reflects estimated coverage issues. The experience from 2001 and 2011 is that the coverage survey demonstrated its capability to estimate and adjust for local council areas with census response rates for the ranges being estimated in Scotland for 2022 Census.

21. However, that has been based on a coverage survey that itself achieves a very high response rate independent of census coverage. NRS has prioritised a strategy to both maximise the overall response in the coverage survey and minimise variation in response. This is important because a high overall response can hide local areas with very low response, and it is these low response areas that become crucial for census coverage.

22. A sub-group of the International Steering Group met weekly with NRS while CCS fieldwork was under way, recommending adaptations such as integrating address listing with interviewing rather than as sequential activities, the movement of temporary field staff between areas with enhancement from professional interviewers, a focus on validating hard refusals where possible with additional follow-up, following-up households where self-response forms were left, and continuing CCS fieldwork in those areas where response was lowest to maximise every household’s opportunity to respond.

23. Despite these efforts, the overall CCS response is just below 60%. This is below target and lower than achieved in 2011, although broadly in keeping with both the ONS 2021 CCS, which achieved just over 60%, and a wider general decline in social survey response rates. While the final overall response rate is lower than achieved in 2011, further continuation would have negatively impacted on the subsequent processing stages while delivering very few additional responses.

24. As a result, it is now necessary for NRS to place much greater reliance on Pillar 3), administrative data, than had been originally planned. The group is providing advice in two broad areas.

a. Exploring the general approach to coverage estimation, recognising that the planned approach based on 2001 and 2011 may not be the best fit for the 2022 situation as it has developed. In particular, the lower coverage survey response points to coverage estimation approaches that facilitate combining across areas, while accounting for individual and household characteristics rather than a strict reliance on geography. This draws on the ONS approach utilised in 2021.
b. Expanding the breadth of Pillar 3) to bring in administrative data earlier and at an individual-level.

Enhancing Pillar 3) to Boost the Value of Administrative Data

25. Administrative data have always been integral to the production of Scotland's 2022 Census official population statistics. Specifically, NRS has planned for the 2022 Census to take advantage of administrative data to provide comparison counts for quality assurance but also allow for direct imputation when a census return is missing key data.

26. The International Steering Group has advised that expanding use of administrative data is a key priority for NRS and we are supporting applications by NRS for enhanced access and use of administrative data throughout the census processing. Specifically, the Group has recommended, and NRS are acting on this, that data linkage of individuals across administrative data, census, and CCS be carried-out using name, date-of-birth, and address information in an appropriate secure setting. This will support the following two objectives; **estimating the size of Scotland's population** and correct placement of imputed records into local areas with lower coverage. The direct use of administrative data in census estimation is already established, for example in Northern Ireland (2011 and 2021) and New Zealand (2018). More generally, international population statistics systems are increasingly comprising composites of conventional enumeration and survey models with data drawn from administrative systems.

27. To meet the first objective, the International Steering Group is working with NRS to enhance their standard estimation procedure by including administrative records in the calculations, either as part of the census or CCS datasets, in place of people who did not respond. Statistical models will then be applied to estimate Scotland's population adjusting for those missing from the census response. Once population estimates have been created, the administrative data will be removed from the datasets prior to the creation of the adjusted census database.

28. The addition of multiple administrative data linked together in the modelling of those missed by the census enumeration is necessary for the production of high-quality population estimates. The ability to augment both census response and CCS response, with the basic demographic characteristics (age, sex, ethnicity) of those identified as missed by administrative data, provides flexibility in the construction of the statistical modelling; and reduces the variation in estimates of census response patterns. As administrative data do not rely on the general public responding, there inclusion strengthens the modelling further by reducing the risk of dependence between census and CCS response. This drives additional improvement in the performance of the statistical modelling reducing bias in the estimates of census response patterns.

29. The second objective is the adjustment of the enumerated data to reflect the estimated coverage errors as the final stage in the production of census outputs.

30. At this point, **enhanced use of individual and household information from administrative data can direct placement** to reflect local areas with problematic coverage. This will directly improve the quality of census information at a local-level by pointing to localised pockets with low census response and guiding the types of individuals and households to be placed by NRS's adjustment system.

31. An important related issue is the measurement of coverage errors for the non-household population. These groups tend to be relatively small in the total population but are important sub-populations that are typically clustered into local areas. Three examples would be:

a. students in halls of residence,
b. the elderly in care homes and other communal settings,
c. prisoners not considered as usual residents of a household (a technical decision based on length of sentence).

32. For these populations, access to specialist administrative data is crucial and a strong focus of the steer from the International Steering Group to NRS has been the importance of these data. NRS has responded to that steer as evidenced by the movement on earlier access to student data.

Concluding Remarks

33. The International Steering Committee consider that **the main census enumeration has provided the foundation for a high-quality set of census outputs, in terms of coverage of the population.**

34. **However, the Census Coverage Survey (CCS) and Administrative Data are essential components of those final outputs.** It is now clear that the CCS alone will not be sufficient to inform the estimation and adjustment processes. Specifically, the expansion and enhancement of administrative data use beyond the original plans for estimation of census response is essential to put NRS in a strong position to deliver a high-quality set of census outputs for Scotland's 2022 Census. With support from the International Steering Group, NRS are working at pace to secure the necessary access to key administrative datasets for the purpose of census estimation and adjustment.

35. Conversely, reliance on the coverage survey alone to deliver the estimation and adjustment will likely result in a final set of outputs that while performing for the high-level counts for Scotland will **not** support statistically robust estimates for all smaller areas and population groups. Therefore, it is clear to us that facilitating NRS to access and utilise the necessary administrative data is central to the final quality of Scotland's 2022 Census.

Annexe B

1. This note provides a very brief summary of the findings of the Call for Views that the Constitution, Europe, External Affairs and Culture Committee launched to inquire into the outcome of Scotland's Census, focusing on the quality of the data collected, the identification of any gaps, and how those could be filled.

2. It follows the Committee's evidence session on 23 June 2022 with the National Records of Scotland.

3. On the closure of the Call for Views on 12 August 2022, the Committee received 27 written submissions – 24 from individuals (including 4 academics and 1 researcher) and 3 from organisations – all of which are available on the Parliament's website.

4. The range of views expressed in the Call for Views covered issues such as—

- How the census was organised, including
    - guidance
    - publicity
    - timing
- Concerns around specific questions, including
    - sex and gender
    - national and ethnic identity
- Quality of the data
- What were seen (by some) as missed opportunities in areas such as
    - volunteering
    - transport
    - neighbourhood initiatives

CEEAC Committee Clerks
August 2022

Annexe C

Census data quality

This note describes the approaches National Records of Scotland (NRS) are using to check and improve the quality of data coming out of the Scottish Census.

Working out how many people are missed from the Census: Census Coverage Survey

To work out credible population estimates, statisticians need to know what proportion of people and households fill in the Census.[1] This is what the Census Coverage Survey is for – similar surveys are run in England and Wales, Northern Ireland and various other countries.[2] Unlike the Census, the Census Coverage Survey is voluntary and goes to only a small fraction of people and households (approximately 1.5-2% of the population), and it asks fewer questions.[3] It ran between 13 June and 22 August.[4]

The methods used to produce population estimates from the Census rely on the assumption that the Census Coverage Survey is equally likely to be filled out by someone who has filled in the Census as someone who has not filled in the Census (and vice versa, within certain groups). Various techniques are used to achieve this statistical independence between the Census itself and the Census Coverage Survey – for example using different methods for the survey. For example, the Coverage Survey is mainly face to face.[5]

In addition, the more people responding to the Census Coverage Survey, the better the results are likely to be. In 2020, NRS said that households will be visited up to 10 times, with a self-response paper questionnaire left on the last visit. The option to fill in the survey by telephone will also be provided from halfway through the fieldwork.[6] Modelling done by the NRS suggested that national targets for census precision would be met with a Census response rate of 94% (in practice it was lower) and a Census Coverage Survey response rate of 80% (in 2011 the return rate was 87%).[7]

No Census Quality Survey

---

[1] National Records of Scotland, Statistical Quality Assurance Strategy, Version 1.0, May 2021

[2] Office for National Statistics, Maximising the quality of Census 2021 population estimates, June 2022; Northern Ireland Statistics and Research Agency, Coverage and quality surveys, undated. Examples of other countries that run similar surveys are the USA and Australia.

[3] National Records of Scotland, Scotland's Census 2021 Estimation and Adjustment Methodology, May 2020; National Records of Scotland, Scotland's Census 2022 Census Coverage Survey Sample Allocation and Reserve Sample Methodology, May 2020

[4] National Records of Scotland, Scotland's Census 2022 - what happens next? updated 23 Aug 2022

[5] National Records of Scotland, Scotland's Census 2021 Estimation and Adjustment Methodology, May 2020

[6] National Records of Scotland, Scotland's Census 2021 Estimation and Adjustment Methodology, May 2020

[7] National Records of Scotland, Scotland's Census 2022 Census Coverage Survey Sample Allocation and Reserve Sample Methodology, May 2020; National Records of Scotland, NRS releases latest return data for Scotland's Census 2022, 9 June 2022

For previous Censuses the Scottish Government have used a Census Quality Survey – the Office for National Statistics still use one for England and Wales. A Census Quality Survey can be used to assess the quality of people's answers to census questions, by looking at how consistent those answers are between the Census and the Census Quality Survey.[8]

For this Census, the NRS will not be running a Census Quality Survey. In May 2021, they explained that:

> Historically the Census Quality Survey (CQS) was a voluntary survey carried out across Scotland eight weeks after the census day. It aimed to capture the differences between responses in Census compared to the CQS.
>
> However, for Scotland's Census 2022 we have changed our approach and we will not carry out a CQS in the traditional way. Instead we will use alternative data quality measures which offer robust and extensive assurance that Scotland's Census 2022 produces high quality data and is of value to the people of Scotland. We have developed additional quality assurance processes that significantly improves our approach to ensure Census data is high quality. The new approach we have adopted relies on the robustness, quality and value of existing statistical quality assurance processes.[9]

An earlier version of the Statistical Quality Assurance Strategy (version 0.5, Nov 2019) seems to imply that a Census Quality Survey was planned at that point.[10] The Census Quality Survey wasn't raised in the public committee session on 23 June.[11]

Alternative approach: using other data to check and adjust the census

As the previous quote says, rather than running a Census Quality Survey, the NRS are now using "alternative data quality measures which offer robust and extensive assurance that Scotland's Census 2022 produces high quality data and is of value to the people of Scotland".[12]

Mostly their quality assurance of particular results will involve the NRS comparing the figures coming out of the Census with statistical counts from other sources – for example the number of children in schools in each area from the Scottish Government Pupil Census, or the number of people in prison from Scottish Government Justice Analytical Services.[13] The NRS say that they will take into account the limitations of these sources – for example the fact that the figures on the

---

[8] National Records of Scotland, Scotland's Census 2022 Statistical Quality Assurance Strategy, May 2021; Office for National Statistics, Census Quality Survey, undated; Scotland's Census 2011 General Report, Oct 2015 – p144 on

[9] National Records of Scotland, Statistical Quality Assurance Strategy, Version 1.0, May 2021

[10] National Records of Scotland, Statistical Quality Assurance Strategy, Version 0.5, Nov 2019

[11] Constitution, Europe, External Affairs and Culture Committee, Meeting date: Thursday, June 23, 2022, Scotland's Census

[12] National Records of Scotland, Scotland's Census 2022 Statistical Quality Assurance Strategy, May 2021

[13] National Records of Scotland, Scotland's Census 2022 Quality Assurance of Administrative Datasets (QAADs), April 2022

number of children from the Pupil Census will not include children in independent schools.[14]

In a couple of cases the NRS will use the records of individual people to check the census data – including records from registration of births, deaths, marriages and civil partnerships registration, and data such as people's names, dates of birth and postcodes from the National Health Service Central Register (NHSCR).[15]

A similar process was carried out following the 2011 Census but used a somewhat different set of data sources used for checking – in general more sources seem to be being used for 2021 but there are a couple of 2011 sources that appear to be missing, for example data from the DWP.[16]

Quality assurance panels will also check results – these will include external members who may have access to further expertise and data for cross-checks, for example local authorities who "may be able to compare estimates to a local data source such as a database of users of a local service or they may have insight into the location and capacity of a new hall of residence or unexpected population movement due to COVID-19."[17]

Amending the estimates

It is not yet clear how much, if at all, the NRS will change census estimates following quality assurance processes.

In 2011, ONS made a 'national adjustment' to the England and Wales census estimates when other sources suggested the census had undercounted the number of males aged 20-49.[18] This was not done in Scotland in 2011 because there was no "robust comparator source" to compare the census to.[19] It is not clear what the planned approach for the 2022 census is.

Comment by the regulator, the Office for Statistics Regulation

The Assessment Programme Lead at the Office for Statistics Regulation recently wrote to the Director of Statistical Services at National Records of Scotland.

He said that he welcomed the work that the NRS had done to improve the quality of Census results:

---

[14] National Records of Scotland, Scotland's Census 2022 Quality Assurance of Administrative Datasets (QAADs), April 2022

[15] National Records of Scotland, Scotland's Census 2022 Quality Assurance of Administrative Datasets (QAADs), April 2022

[16] National Records of Scotland, Scotland's Census 2022 Quality Assurance of Administrative Datasets (QAADs), April 2022; National Records of Scotland, Scotland's Census 2022 Validation of Population Estimates Methodology, November 2021; Overview of Administrative Comparator Data Used in 2011 Census Quality Assurance, November 2012

[17] National Records of Scotland, Scotland's Census 2022 Statistical Quality Assurance – Validation of Population Estimates Quality Assurance Panels, August 2022

[18] National Records of Scotland, 2011 Census Release 1B - How the 2011 Census population estimates were obtained, March 2013

[19] National Records of Scotland, Scotland's Census 2021 Estimation and Adjustment Methodology, May 2020

> The disruption caused by the COVID-19 pandemic and the change in both timing and mode of data collection to digital first mean that the context of this census is noticeably different from previous ones. There has been much coverage recently of the level of response and concerns raised about the possible impact of this on the quality of population estimates from the census. We share your concerns that the focus on the census response rate as the sole indicator of the quality of estimates could become a distraction. In particular, we consider that questions about what response rate is sufficient do not adequately capture that population outputs do not rely on census data alone, but a combination of census and other data.
>
> Having said that, we welcome the steps taken by NRS to achieve the best response rate possible and to maximise the usefulness of the census coverage survey. The additional steps that NRS is taking, such as investigating what additional non-census data can be used to further improve the quality of census estimates and establishing an International Steering Group to provide additional expert input, all point towards an endeavour that continues the aim of producing the best census outputs possible.

However, he also suggested that the NRS should be publishing more about what it is doing to improve data quality:

> It will […] be important for NRS to update the existing materials soon to reflect how you are adapting your processes, and the possible impact on the expected quality of census estimates. We are mindful of the need to be proportionate in doing this, especially when the focus of your team is on processing census data. At the very least we expect a clear description of how the various steps you are taking help to assure the quality of census estimates fit together, and to provide updates to users about the level of quality that they may expect from census outputs.[20]

SPICe

30 August 2022

[20] Letter from Mark Pont to Pete Whitehouse: Scottish Population Census, 17 August 2022