Dear Convener

**SCOTLAND’S CENSUS 2022**

During the Committee session on the 29th October 2020 to discuss the draft Census (Scotland) Amendment Order 2020, I agreed to provide further information on how National Records of Scotland will use the Census returns to produce our Census Output estimates. This was in response to a question from Kenneth Gibson MSP. I hope the following is of interest and use.

The starting point for all Censuses is to maximise the number of individuals and households who provide a complete Census form. This is typically done by:

- Developing questions which are able to be understood and answered by the whole population
- Providing support and advice to help forms to be completed – be this translation services, a customer helpline, guidance accompanying the form, local support within communities etc
- Ensuring accessibility to enable and support disabled people to participate, both through the use of supportive digital technology and through our service design
- Promoting awareness and importance of the Census through media, advertising and promotional activities
- Providing options for completion such as online, paper completion, and telephone data capture
- Employing a field force to encourage and support census completion by individuals and within local areas and communities.

The forms returned are then taken through a series of data cleansing¹ and processing procedures which help to assure and where necessary improve the quality of the data. Changes and corrections that might be made could include where a child aged under 16 reports working in full time employment, a birth date being before 1900, or a pupil under 17 reports driving themselves to school.

¹ Data cleansing is the process or detecting and correcting corrupt or inaccurate records
Through the combination of editing\(^2\) and imputation\(^3\) work, complete and consistent Census returns are created as the basis for producing high quality Census Outputs. This work has been a standard and important element of Census work since 1981.

Whilst the ambition of all Censuses is to maximise the number of respondents, it is recognised that not everyone will respond. In Scotland the Census is underpinned by a legal responsibility. Where a householder refuses to complete a census questionnaire non-compliance communications will focus on maximising response. In a small number of cases, a householder may be referred for consideration of prosecution by COPFS. However, this is a potentially lengthy process and therefore statistical techniques to estimate for the level and impact of non-response are employed.

Scotland’s Census in 2011 saw a 94% response rate of the population, which is similar to Censuses across the world [England and Wales (2011) – 94%; Northern Ireland 94%; Australia (2016) – 94.8%]. The objective for Scotland’s Census in 2022 is that response rates will be similar to 2011.

The challenge for Census producers (across the world) is therefore how to understand who has been missed by the Census and then how to include this information (to both count the missing individuals and to provide information about their characteristics and likely response to Census questions). We need to estimate these missing individuals to ensure that when the results are being used to report statistics they are representative of the ‘whole’ population and that resources are allocated fairly, based on estimates for the ‘whole’ population rather than a potentially unbalanced subset.

The first stage is to produce an estimate of the total population. This is done by making use of the Census Coverage Survey (CCS). This secondary voluntary survey runs shortly after the main census collection, involves trained interviewers who collect data mainly through face to face interviews with around 40,000 households. It takes place 6 weeks after the Census date, and gathers information which helps to estimate the number of households and individuals who were missed by the Census. This is done by using the field force to gather information on households and individuals in a given area, and then comparing this information with the population figures gathered by the Census. Comparing these two population figures then creates an adjustment factor which is used to produce the true estimated population total. Once the ‘true’ population figure is calculated, Census returns are created for those individuals and households who are considered to be missing from the initial Census returns. These ‘synthetic’ returns are created using evidence from the Census to ensure that the new data are plausible – this is done by exploring the patterns and links between individual characteristics and question responses captured in the Census and producing likely responses for the missing individuals.

Therefore the ‘complete’ dataset, which in 2011 consisted of 94% Census returns plus a further 6% where missing data was estimated for, is created and is then used as the basis for all Census Outputs.

\(^2\) Data editing is a process through which data is reviewed to check for consistency, errors etc

\(^3\) Data imputation is a process of substituting estimated values for missing data
The approach taken in Scotland, and which will be taken in 2022, is similar to that used across the rest of the UK. It is an approach that has been reviewed by independent experts and is recognised as an important component in the delivery of high quality Census Outputs which represent the complete population of the country.

I hope that you find this information to be helpful.

Yours sincerely

Pete Whitehouse
Director of Statistical Services