

A negligible effect? The untested assumptions behind plans for the 2021 census

Dr Kath Murray
Lucy Hunter Blackburn
Lisa Mackenzie

4 June 2020

www.murrayblackburnmackenzie.org

“The proposed 2021 census will probably be the most important for a generation. It will begin to show the devastating effects of Covid-19 on our communities and individuals. It is imperative that the largest number of our citizens are included.”
(Lord Goddard, [House of Lords debate](#) col. 585, 12 May 2020)

“The planning and execution of the decennial population census is a vast and complex undertaking often described as the largest peacetime operation carried out in the UK. It is certainly the largest statistical exercise, aiming to collect socio-demographic information from every individual and household in the four home nations.”
(Compton et al. 2017: 33)

The 2021 UK census will take place in just under 10 months, on 21 March 2021. In England, Wales and Scotland, the questionnaire will include two new, voluntary questions about sexual orientation and gender identity, as well as the longstanding sex question, which is compulsory.

As the Office for National Statistics [explains](#), sex is a key demographic variable and robust data is vital to census users:

“sex is taken into account in a huge proportion of the analysis that goes on across government, as well as in academia and other spheres. In many cases this will be in examining any differences in results by sex, while in other cases sex is used as a weighting variable (to allow more accurate comparisons between different groups to be made). Data on sex is of key importance for equality monitoring.”

The sex question will simply ask whether a person is male or female. However, the three UK census authorities (Office for National Statistics [ONS], National Records of Scotland [NRS] and Northern Ireland Statistics and Research Agency [NISRA]) also intend to accompany the question with [guidance](#) that advises respondents that they can answer based on their self-declared gender identity.

The proposed guidance, which deliberately conflates two separate demographic characteristics, has raised serious disquiet among relevant experts. Strikingly, NRS has also [previously stated](#) that its intention ‘has never been to conflate sex and gender identity’.

In September 2019 a group of social scientists with expertise in quantitative analysis [wrote](#) to the Scottish Parliament Culture, Tourism, Europe and External Affairs (CTEEA) Committee, stating that the proposed guidance would “*reduce the ability of the Census and these other sources to distinguish the situation of those who are male and female, and hence to capture sex-based discrimination and disadvantage*”.

In December 2019, 80 of the UK’s most eminent social scientists [wrote](#) to the three census authorities, setting out their concerns, in particular about the potential impact on data reliability at the subgroup level:

“It is unlikely that the trans population will be evenly distributed across the population, for example by age, sex and geography. This means that the effects on data reliability are likely to be greater at the sub-group level. This can have extreme consequences for particular subgroups, e.g. 1 in 50 male prisoners in England and Wales identify as transgender. The Tavistock and Portman NHS Trust claims that between 1.2% and 2.7% of children and young people are ‘gender-diverse’.”

Such concerns are supported by a [Swedish study](#) based on a population-representative sample of 50,157 Stockholm County residents aged 22 and older (Åhs et al. 2018). The study found that amongst young people aged 22 to 29 years, 4% identified as members of the opposite sex, compared to a 2.3% sample average, and that 6.3% wanted to be treated as members of the opposite sex, compared to a 2.8% sample average. The study also reported that a higher proportion of females wanted to be treated as a member of the opposite sex, compared to males, at 3.5% and 2.0% respectively. These results are shown in **Table 1**.

Table 1. Proportion of adults reporting other aspects of gender incongruence, by sex and age

	I feel like someone of a different sex		I would like to live as or be treated as someone of a different sex.	
	<i>n</i>	Weighted % (CI)	<i>n</i>	Weighted % (CI)
All	770	2.3 (2.1–2.6)	779	2.8 (2.4–3.1)
Registered Sex				
Male	309	2.1 (1.7–2.5)	218	2.0 (1.6–2.5)
Female	461	2.5 (2.1–2.8)	561	3.5 (3.0–3.9)
Age group				
22–29 years	72	4.0 (2.8–5.2)	107	6.3 (4.7–7.9)
30–44 years	219	2.5 (2.1–3.0)	284	2.9 (2.5–3.4)
45–66 years	329	2.0 (1.7–2.3)	286	2.0 (1.7–2.3)
67+ years	150	1.1 (0.9–1.4)	102	1.0 (0.6–1.1)

N = 50,157

Source: Åhs et al. 2018

As we [wrote](#) earlier this year, in Belgium, levels of applications to change legal sex from female to male amongst those aged 16-24 years old were around five times higher than would be expected from the population share of women that age.

It is not difficult to see how the data's integrity might be damaged at the subgroup level by the proposed framing of the sex question as a self-declared gender identity question. Such effects are likely to be further exacerbated when sex and age are cross-referenced with other variables. While Åhs et al. do not provide an age/sex breakdown, from the data made available it can be reasonably assumed that the proportion of females aged 22 to 29 who wish to be treated as the opposite sex will be even higher than the age-group average (6.3%).

Despite increasing concerns about data reliability, as far as we have been able to establish, no organisation which supports the inclusion of self-identification guidance – from the census authorities to the coalition of Scottish women's groups¹ that submitted [evidence](#) on the Census Amendment (Scotland) Bill – has demonstrated that they have considered the potential impact on data reliability at the population subgroup level, or the implications for and the quality of analysis that will be possible using this data.

On 30 January 2020, when [giving evidence](#) to a committee of the Scottish Parliament, representatives of NRS [failed to respond](#) to a question from Convener Joan McAlpine MSP about the impact on subgroup population data.

At an [earlier evidence session](#) in September 2019, NRS' then director of statistical services Amy Wilson acknowledged that despite similar guidance being introduced in the 2011 census (albeit without any democratic oversight or wider scrutiny) the effect on data quality remained unclear:

"I do not think that we know how it affected the data in 2011. From looking at the data and the quality assurance that we have done, there is no evidence to suggest that we started to see trends that were different from anything that had happened in the past. However, you are right—we do not know how the guidance affected people and we do not know how many people actually looked at it in 2011."

Professor Sullivan and her co-signatories did not receive a response to their letter from the census authorities until 26 February 2020, nor did the response address their substantive concerns. While they indicated that they would provide an opportunity for the signatories to engage in further discussions, it is our understanding that almost six months after they raised their concerns no such meeting has yet taken place. Professor Sullivan has also [written](#) about her experiences of raising concerns with the census authorities, and how this led to her being no-platformed by social research organisation NatCen.² (3).

In April 2020 we submitted Freedom of Information requests to ONS, NRS and NISRA, asking what analysis they had undertaken to estimate the impact on subpopulation data

¹ We wrote to Engender, the feminist advocacy group that co-ordinated the submission, to ask for sight of the statistical modelling they were drawing on to support their assertion that the collection of sex disaggregated data based on self-declared gender identity was unproblematic and the impact on data quality negligible. They responded that they did not have anything to add to their published briefings on this issue.

² In 2019, NRS commissioned ScotCen, an integral part of NatCen, to undertake additional testing on the sex and gender identity census questions.

reliability. The responses received from [ONS](#) and [NRS](#) indicate that neither has undertaken any such analysis. [NISRA](#) did not respond to the FOI, citing pressures on their workload relating to the Covid-19 pandemic.

“We have not undertaken work on the guidance with a specific focus on the quality of analyses possible using 2021 Census data for specific groups.” [ONS]

“National Records of Scotland does not have the information you have requested.”

The proposed self-identification guidance also risks an increase in the number of people who refuse to answer the sex question. While the sex question usually has an extremely high response rate,³ [research](#) commissioned by NRS and undertaken by ScotCen in 2019 revealed that 3% of the general population sample said they would not complete the census if guided to answer based on either their legal (as opposed to biological) sex or self-declared gender identity.⁴

Census data matters. It matters because it drives public spending priorities and decisions about resource allocation. Introducing the recent House of Lords debate on the Census (England and Wales) Order 2020, Cabinet Office Minister Lord True stated:

“The census is the most important source of statistics about the UK population available to us. It is currently the only data collection exercise that provides accurate and reliable information about populations at a local area level. It provides underlying information needed to inform a wide range of policy decisions, and it is used extensively to plan services and allocate funds to local areas.”

Lord True, [House of Lords debate](#) col. 578, 12 May 2020

Census data also acts as the denominator for calculating disease prevalence rates and prevalence rates of other socio-economic phenomena. There is, for instance, a pressing need for robust, high quality data which is disaggregated by sex, age, ethnicity and other key demographic variables on outcomes related to Covid-19.

The UK [Code of Practice for Statistics](#) states that,

“statistics have to be based on the right data sources, with transparent judgements about definitions and methods, and judgements about the strengths and limitations of the statistics. Producers should demonstrate how they assure themselves that their statistics are robust and reliable.”

We believe that the material above strongly suggests a failure by the census authorities to properly engage with and investigate the data reliability concerns raised by relevant data experts and that this represents a departure from the requirements of the Code which needs to be addressed urgently, particularly in light of the data from Sweden. We also question the assumption that clearly defined data on sex is no longer required or relevant for the purposes of the UK’s most important data collection exercise, which appears to have been made without any proper consideration and despite strong statements about the importance of sex made by the census authorities themselves.

³ In 2011, according to the NRS, the non-response rate for the sex question was 0.8%.

⁴ ScotCen tested two versions of guidance for the sex question: one which asked for respondents’ legal sex and one which asked for respondents’ self-declared gender identity.

References

Åhs, J. W., Dhejne, C., Magnusson, C., Dal, H., Lundin, A., Arver, S., Dalman, C., & Kosidou, K. (2018). 'Proportion of adults in the general population of Stockholm County who want gender-affirming medical treatment'. *PloS one*, 13(10) e0204606.

Compton, G., Wilson, A. and French, B. (2017) 'The 2011 Census' in (eds) Rees, P. and Stillwell, J. *The Routledge Handbook of Census Resources, Methods and Applications, Unlocking the UK 2011 Census*. Routledge