CPC
CENTRE FOR POPULATION CHANGE

CONNECTING GENERATIONS

Estimating the 2011 total fertility rate for England & Wales and Scotland using alternative data sources

106 WORKING PAPER SEP 2023

Joanne Ellison
Bernice Kuang
Sarah Christison
Ann Berrington
Hill Kulu

Producing novel science
Informing policy
Improving knowledge on population change
cpc.ac.uk
ISSN 2042-4116

UKRI
Economic and Social Research Council

ABSTRACT

It is important to estimate fertility rates as accurately as possible in order to make appropriate comparisons of fertility levels across time and space and to inform fertility projections. This paper compares estimates of the 2011 total fertility rate (TFR) for all, UK-born and non-UK-born women in England & Wales and Scotland, obtained using several data sources. The three data sources we use are vital registration (VR) data, longitudinal studies (linked census and vital events data) and census household microdata samples from the respective countries. Although estimates based on VR data are classed as official, the event and risk population information come from different sources. Surveys and census data do not suffer from this issue, but their analysis requires decisions regarding the selection of the sample and how to deal with exits and entries to the UK. We find:

- TFR estimates from the census microdata tend to be closest to those from VR data, particularly for Scotland. For England & Wales, the census estimates are lower than those from VR data, especially for non-UK-born women.
- The longitudinal study estimates are the lowest among the three data sources for Scotland, while for England & Wales they are lower or higher than the corresponding VR estimate with this generally depending on the precise estimation method used.
- Overall, this study finds some small variation in the TFR estimates from these different sources, owing to their contrasting coverage, mode of collection and sample size.
- The reasonable consistency of the census-linked data and the census household microdata with the VR estimates shows that they are an important source of information which allows the examination of subgroup differences in childbearing behaviour.

KEYWORDS

Total fertility rate; census; longitudinal study; vital registration; England; Wales; Scotland.

EDITORIAL NOTE

Joanne Ellison is a Research Fellow at the University of Southampton, working within the modelling strand of the ESRC Centre for Population Change Connecting Generations.

Bernice Kuang is a Research Fellow at the University of Southampton, working on fertility and family change.

Sarah Christison is a Research Fellow at the University of St Andrews, working within the fertility and family strand of the ESRC Centre for Population Change.

Ann Berrington is a Professor of Demography and Social Statistics at the University of Southampton and jointly co-ordinates the fertility and family strand of the ESRC Centre for Population Change.

Hill Kulu is a Professor of Human Geography and Demography at the University of St Andrews, and Co-Director of the ESRC Centre for Population Change Connecting Generations.

Corresponding author: Joanne Ellison, J.V.Ellison@soton.ac.uk

ACKNOWLEDGEMENTS

This work was supported by the ESRC awards: Understanding Recent Fertility Trends in the UK and Improving Methodologies for Fertility Forecasting (Grant ES/S009477/1) and the Centre for Population Change Connecting Generations Centre (Grant ES/W002116/1). For further details about the wider project, please go to https://fertilitytrends.wp.st-andrews.ac.uk/.

The permission of the Office for National Statistics (ONS) to use the Longitudinal Study (LS) is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). CeLSIUS is funded by the ESRC under project ES/V003488/1. The authors alone are responsible for the interpretation of the data. This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

The help provided by staff of the Longitudinal Studies Centre – Scotland (LSCS) is acknowledged. The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. The authors alone are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the King's Printer for Scotland.

The authors are grateful to the ONS Secure Research Service and National Records of Scotland for providing access to and clearance for outputs from our analyses of the ONS LS, the SLS, and the 2011 Census Secure Household Microdata samples for England & Wales and Scotland.

ESRC Centre for Population Change
Connecting Generations

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews and Stirling, in partnership with the Office for National Statistics (ONS) and the National Records of Scotland (NRS).

Connecting Generations (CG) is a major strategic collaboration between CPC, the Leverhulme Centre for Demographic Research (LCDM) at the University of Oxford, and the Resolution Foundation.

The Centre is funded by the Economic and Social Research Council (ESRC) grant numbers RES-625-28-0001, ES/K007394/1, ES/R009139/1 and ES/W002116/1.

This working paper series publishes independent research, not always funded through the Centre. The views and opinions expressed by authors do not necessarily reflect those of the CPC, CG, ESRC, ONS or NRS.

The ESRC Centre for Population Change Working Paper Series is edited by Teresa McGowan; t.mcgowan@southampton.ac.uk

ESTIMATING THE 2011 TOTAL FERTILITY RATE FOR ENGLAND & WALES AND SCOTLAND USING ALTERNATIVE DATA SOURCES

TABLE OF CONTENTS

1. INTRODUCTION

It is important to estimate fertility rates as accurately as possible in order to make appropriate comparisons of fertility levels across time and space (Hellstrand et al. 2020; Campisi et al. 2020), and to inform fertility projections (Raftery et al. 2012). The total fertility rate (TFR) is the most commonly used summary fertility measure (Hinde 2014). The TFR is an indicator of the fertility level in a particular country or region; it measures the average number of children a woman would have if she were to experience the age-specific fertility rates (ASFRs) of a given period throughout her reproductive career (Ellison et al. 2020). It is calculated by summing the ASFRs (weighted by the corresponding width of the age group if single years of age are not used) across the reproductive age range for the period of interest, typically a calendar year (Jasilioniene et al. 2015). In this paper we compare estimates of the 2011 TFR for England & Wales (treated as a single entity) and Scotland, obtained using several data sources which differ greatly in their coverage, mode of collection and sample size.

The data requirements for computing an ASFR for a given age and period are the number of live births to women of that age during that period (numerator) and the corresponding number of person-years lived by the female population (denominator). For developed countries, official estimates produced by national statistical agencies (NSAs) source the numerators from high-quality vital registration (VR) data and approximate the denominators using census- or register-based population estimates (United Nations Department of Economic and Social Affairs 2007; WHO 2023). Sample surveys tend to be used to compute the TFR when VR data of a sufficient level of completeness is not available (United Nations Department of Economic and Social Affairs 2022). Such alternative data sources provide the opportunity to also carry out in-depth fertility studies, for example by including additional characteristics such as education or ethnicity or tracing individuals over time. Therefore, it is of interest to compare the estimates of simple fertility indicators such as the TFR from these sources with the estimates from VR data for the purposes of validation, calibration and assessing representativeness[1].

For each country, we compute estimates of the 2011 TFR from three data sources: a) VR data; b) the Office for National Statistics (ONS) and Scottish Longitudinal Studies (ONS LS and

[1] Survey and census data can also be used to fill in gaps or adjust VR data, e.g. the Office for National Statistics (ONS) use the Annual Population Survey (APS) to estimate the female population by age and country of birth when calculating estimated TFRs for UK and non-UK-born women (ONS 2022a); every 10 years, ONS also revise their mid-year population estimates based on population counts from the most recent census (ONS 2022b).

SLS); and c) the respective 2011 Census Secure Household Microdata samples. Although the UK has high-quality VR data to inform the ASFR numerators, there is greater uncertainty regarding the accuracy of the denominators, which are taken from a different source. This highlights a disadvantage of the construction of the VR estimates, namely that they do not adhere to the principle of correspondence: this stipulates that, for a given rate, the events in the numerator and the population exposed in the denominator should always correspond with each other (Hinde 2014). The longitudinal studies overcome this drawback by linking the same VR data to census data for a random sample of the population; however, it is necessary to make decisions as to which women should be included, and how entries and exits between censuses should be handled. Lastly, the census microdata samples are larger in size compared to the longitudinal studies and include the events and risk population within the same source, which is beneficial; sample selection decisions still need to be made though, and there can be difficulties when linking mothers with babies. In summary, there is no 'gold standard' method for estimating fertility rates, with each of the data sources we consider here presenting different strengths and challenges.

In 2011, the TFR was 1.93 in England & Wales and 1.69 in Scotland (ONS 2023d). Significant numbers of births occurred to non-UK-born women: they accounted for 25.5% and 14.2% of the total births in England & Wales and Scotland respectively (NRS 2012; ONS 2012). The corresponding estimated TFR for non-UK-born women in England & Wales was 2.21 (compared to 1.86 for UK-born women), and in Scotland was 1.91 (compared to 1.69 for UK-born women) (ONS 2019b, 2022a). The share of the immigrant population is increasing, with 28.8% and 17.3% of births in England & Wales and Scotland respectively occurring to non-UK-born women in 2021 (NRS 2022; ONS 2022a). It is therefore of interest to compare TFR estimates from the alternative data sources (longitudinal studies and census microdata) for these population subgroups, i.e. UK-born and non-UK-born women, as well as overall for all women.

The paper proceeds as follows. In Section 2 we provide further information on the data sources, sample selection and computation methods. We then present the England & Wales and Scotland 2011 TFR estimates calculated from the three data sources in Section 3, for the different country of birth categories. Finally, we conclude with a discussion in Section 4.

2. DATA AND METHODS

2.1 ESTIMATED FERTILITY RATES FROM VITAL REGISTRATION

The NSAs of England & Wales and Scotland (ONS and National Records of Scotland, NRS) take a similar approach to the calculation of official birth rates, whereby the birth registration data informs the numerators of the ASFRs[2]. Births are only included if they take place within the respective countries, whether they are to residents or visitors (NRS 2023c; ONS 2023c). However, births to residents that take place outside of the particular country are excluded. Mid-year population estimates of the usually resident population are used to inform the denominators of the ASFRs (NRS 2023d; ONS 2023c)[3]. It is important to note that this method goes against the principle of correspondence (Section 1) as, for example, the numerator could include births to women no longer present at the mid-year and therefore absent from the denominator; alternatively, a woman could contribute to the numerator and denominator of different single-year ASFRs if she was age x at the mid-year but had a birth following this at age $x + 1$. Despite this drawback, mid-year population estimates provide a reasonably close approximation to the exposure time which improves with population size (Hinde 2014). The ASFRs are calculated for single years of age, with the reproductive age range assumed to be 15-44 and births occurring at ages below 15 and above 44 included in the 15 and 44 age categories respectively (NRS 2023a; ONS 2023c).

We source the 2010 and 2011 birth registration data from ONS (2023a) and NRS (2022), and the corresponding mid-year population estimates from ONS (2021a). We also source estimates of ASFRs for UK-born and non-UK-born women living in England & Wales (ONS 2022a) and Scotland (ONS 2019b). To improve comparability with the TFR estimates based on infants present in the 2011 Census, we compute the VR estimates as weighted averages of the 2010 and 2011 data to more accurately represent the year to census day.

---

[2] One difference is that while ONS consider births by date of occurrence and allow for late registrations of births that occurred in the year of interest, NRS consider births by date of registration (NRS 2023b; ONS 2023c).
[3] Mid-year population estimates for the years between two population censuses are obtained by adding annually information on births, deaths and migration (emigration and immigration) to census data, with information on international migration estimated from (passenger) survey data (ONS 2021b).

2.2 ESTIMATING FERTILITY RATES USING THE CENSUS LONGITUDINAL STUDIES

The longitudinal studies consist of 1% (ONS LS) and 5.3% (SLS) samples of the England & Wales and Scotland populations respectively, for whom census and vital events data are linked (ONS 2023b; SLS Development & Support Unit (SLS-DSU) 2023). The England & Wales population sample that the ONS LS consists of is selected using four birth dates, while 20 birth dates are used for the Scottish population sample that makes up the SLS. The ONS LS links census data from 1971, 1981, …, 2011 with vital events registration data up to 2017, while the SLS links census data from 1991, 2001 and 2011. Among other uses, the LS data enable detailed analysis of period and cohort fertility trends by age and parity, i.e. number of previous children, with sample sizes adequate for performing subgroup analyses, e.g. by educational level or ethnicity. The ONS LS (ONS 2019c) is accessed via the ONS Secure Research Service (SRS), and the SLS (SLS-DSU 2023) is accessed via the NRS Safe Setting.

2.2.1. EVENTS AND EXPOSURES

The quality of the linkage of birth registrations to ONS LS sample mothers has increased over time and is now at a very high level (ONS 2019a). In contrast, the computation of the denominators presents many challenges. While it can be reasonably assumed that a LS member is present in the country during a census year based on census presence, in between censuses it is necessary to use GP registrations to provide information on entries, exits and re-entries, which may be incomplete or inconsistent, with each other, or with census presence. However, even just focusing on consistent cases, previous studies have shown that fertility patterns differ greatly between those LS members continuously and non-continuously resident between censuses, and that including all consistent cases improves the correspondence with VR estimates (Robards et al. 2011). This highlights the additional complications involved in using LS data for fertility analyses, and the need to think carefully about who is included in the sample.

2.2.2. METHODS FOR CENSUS LONGITUDINAL STUDIES

We use two methods to calculate the 2011 TFR using the longitudinal studies – a longitudinal method and a cross-sectional method. The longitudinal method is applied for all years, with births as the event of interest and exposure time assessed using a combination of data on census presence and migration events. The cross-sectional method is a calculation done only at census years, taking the number of births that occurred in the census year and dividing it by the number of women present at the census for that year. We use this approach because census presence would be able to tell us clearly about exposure time for a census year. The main difference between the two methods is that for the longitudinal method, the events and risk time always come from the same people, while the cross-sectional method is analogous to the VR approach (Section 2.1) in that the numerator and denominator essentially come from different sources (birth registration data and census data). We describe the sample selection processes under the two methods for the ONS LS and SLS in Sections 2.2.3 and 2.2.4 respectively.

2.2.3. ENGLAND & WALES SAMPLE (ONS LS)

We start with the same dataset for both methods, consisting of 439,778 women. However, the number of women ultimately used to determine the exposure time differs between methods. The sample used for the longitudinal method drops women with no census or trace information, women who embarked before age 15 and did not re-enter via a re-entry event, women who died before age 15, women who immigrated into the country after age 50, and women who were lost to follow-up before age 15 (i.e., disappear from the census as children) and did not re-enter.

Dropping women from the sample means that any events associated with them are also dropped. Moreover, for any given year, if the woman's presence is not confirmed via census (within five years), then she is also dropped at that point in time because we assume that she is not in the country – this means that any birth event associated with her around that time frame is also dropped. For example, someone present for the 1991 census but absent for the 2001 census will be dropped from the risk set halfway through the 10-year period, at 1996, even if she gave birth in 1997. We do this because including her until 1997 would inflate and bias exposure time for those who give birth.

The sample used for the cross-sectional method does not drop anyone at the outset, but does require that women be present at census to be in the denominator for the TFR in a census year. We do not exclude any births, so births may be to women present or absent at the time of the census. This is similar to the VR approach (Section 2.1), where all registered births (that is, all births that take place in England & Wales) are included in the event count and exposures are determined using the census-based mid-year population estimate.

2.2.4. SCOTLAND SAMPLE (SLS)

The Scotland sample selection process is very similar to that for England & Wales. The initial dataset consists of 151,850 women. The sample used for the longitudinal method drops women with no census or trace information, women with low quality date of birth data, women who experienced births before the age of 14, women who left Scotland before the age of 16 and did not re-enter, and women who died before the age of 16. Women who immigrated into Scotland and were present for some portion of the reproductive age range (ages 16-50) were included in the sample, with episodes before migration dropped from the analysis; these women therefore only began to contribute person-time following their migration. Women whose trace information flagged them as immigrants but had no date of migration were dropped from the sample. As with the approach used in England & Wales, re-entries were permitted.

2.3 ESTIMATING FERTILITY RATES USING CENSUS HOUSEHOLD MICRODATA

The census microdata consists of responses from the residents within a nationally representative sample of 10% of households for the respective countries (ONS 2016a). It provides a rich source of fertility information at a particular point in time through its large sample size, high coverage of the population and household grid structure, which describes the relationships between a given household member and all other household members (e.g. parent, partner, child, sibling). This information can therefore be used to associate mothers with the children who are living with them and their ages on census day. As our interest is in estimating the TFR for the year to census day (27 March 2011), i.e. roughly 1 April 2010 – 31 March 2011, our aim is to link babies aged under 1 at the time of the census with their mothers. The accuracy of this approach relies on a high proportion of these babies living with their mothers, and, where this is the case, a high proportion of the mother-child relationships being accurately reported and recorded in the census. The 2011 Census Secure Household Microdata samples for England & Wales (ONS 2016b) and Scotland (NRS 2016) are accessed via the ONS SRS.

2.3.1. METHOD FOR CENSUS HOUSEHOLD MICRODATA

We describe the method firstly for England & Wales. We begin by considering the babies (individuals aged 0) in the sample, and try to link as many of them as possible to their mothers if they are living in the same household. In the vast majority of cases, one mother is consistently indicated through the household grid[4]. In the remaining cases we concentrate on the particular

[4] Where two mothers are consistently indicated through the household grid, we assign the baby to the older mother.

family unit that the baby belongs to within the household. We make use of additional variables in an attempt to identify the mother, starting with derived fertility variables and then moving on to contextual variables relating to family type. We remove any babies who cannot be confidently linked to a mother, as well as those with mothers aged outside of the standard reproductive age range (15-49). Linkage difficulties are the predominant reason for removal, with the issue much more likely to be that no potential mother figure is identifiable within the family unit rather than there being multiple candidates.

Next, we consider all women of reproductive age. We remove students or schoolchildren living away during term time. We also exclude women (and any babies linked to them) who were living outside of the country of interest one year prior to the census, and therefore contribute less than one full year of exposure time[5]. Then, for each birth, we adjust the age of the mother at the time of the birth from her current age to a year younger where necessary, i.e. where the mother's birthday falls between the time of the birth and census day; we impute the baby and/or mother month of birth if missing. For each woman, we also apportion the year of exposure time between the current age and previous age using her month of birth. Lastly, we aggregate the births across 5-year age groups of mother's age, and the exposure time across 5-year age groups of woman's age. Dividing the former by the latter gives the ASFRs, which we sum to compute the TFR estimate.

We take an almost identical approach for Scotland. However, one key difference is that only the relationship of each household member to the household reference person and their spouse is available, rather than the relationships to all other household members which we have for England & Wales. We therefore need to adjust our method slightly. We give an overview of the sample selection process for each country in Sections 2.3.2 and 2.3.3.

2.3.2. ENGLAND & WALES MICRODATA SAMPLE

There are 71,457 babies in the England & Wales microdata sample. Of these, 67,213 (94.1%) have one mother consistently indicated through the household grid, and 3,283 (4.6%) are removed due to not being able to confidently identify a mother or having a mother aged outside of the reproductive age range. There are 1,391,617 women of reproductive age in the sample, 29,691 (2.1%) of whom are students or schoolchildren living away during term time and are

[5] We have investigated the sensitivity of the TFR estimates to this exclusion. The results are very similar overall and for UK-born women, but are noticeably lower for non-UK-born women.

removed. Of the remaining women, 1,336,046 (98.1%) were living in England & Wales one year prior to the 2011 Census and are therefore included in our final sample.

2.3.3. SCOTLAND MICRODATA SAMPLE

There are 6,029 babies in the Scotland microdata sample. Of these, 5,364 (89.0%) have one mother consistently indicated through the less detailed household grid, and 170 (2.8%) are removed. There are 131,018 women of reproductive age, 2,043 (1.6%) of whom are students or schoolchildren living away from home and are removed. Of the remaining women, 126,039 (97.7%) were living in Scotland one year prior to the census are included in our final sample.

3. RESULTS

In Figure 1 we plot the 2011 TFR estimates computed from the three data sources as described in Section 2. The top and bottom rows present the estimates for England & Wales and Scotland respectively, while the first, second and third columns contain the estimates for all, UK-born and non-UK-born women respectively. The horizontal lines are at the level of the corresponding VR estimate to make it easier to compare estimates across data sources and methods. As discussed briefly at the end of Section 1, the VR estimate for UK-born women is substantially lower than that for non-UK-born women for both countries; this is supported by the estimates from the alternative data sources. In contrast, the overall TFR estimates are very close to those for UK-born women, which is not surprising given that this group makes up the vast majority of the total population (in 2011 the proportion was estimated to be 86.9% and 93.5% for England & Wales and Scotland respectively (ONS 2017)).

Taking the estimates from the census microdata first, we see that for England & Wales they are all lower than the corresponding VR estimates, the greatest relative decrease being observed for the non-UK-born women (11.5%, compared to 4.1% and 2.7% for all and UK-born women respectively). For Scotland, however, the estimates are much closer overall, exhibiting 1.8% and 0.6% increases for all and UK-born women, and a 1.0% decrease for non-UK-born women. This is likely to be at least partially due to the larger proportion of babies included in the census microdata sample for Scotland compared to England & Wales (97.2% vs 95.4%; see Sections 2.2.3-2.2.4). We consider this in more detail later in this section.

Moving to the LS estimates, we note that for UK-born women the cross-sectional estimate for Scotland is not available, and for non-UK-born women only the cross-sectional estimate for

England & Wales is available. Where the estimates from both methods are available, we see that the cross-sectional estimate is considerably higher than that for the longitudinal method. The cross-sectional method includes more births than the longitudinal method, which is expected because no events are dropped when we conduct the cross-sectional calculations, and for 2011 it also includes more exposure time. Therefore, the higher cross-sectional estimates are as a result of the numerator increasing by a larger relative magnitude compared to the denominator. Contrasting the LS estimates with those from VR, we see that for England & Wales the longitudinal and cross-sectional estimates are lower and higher respectively, by similar relative amounts for all women and UK-born women (4.1% and 3.6% for all women,

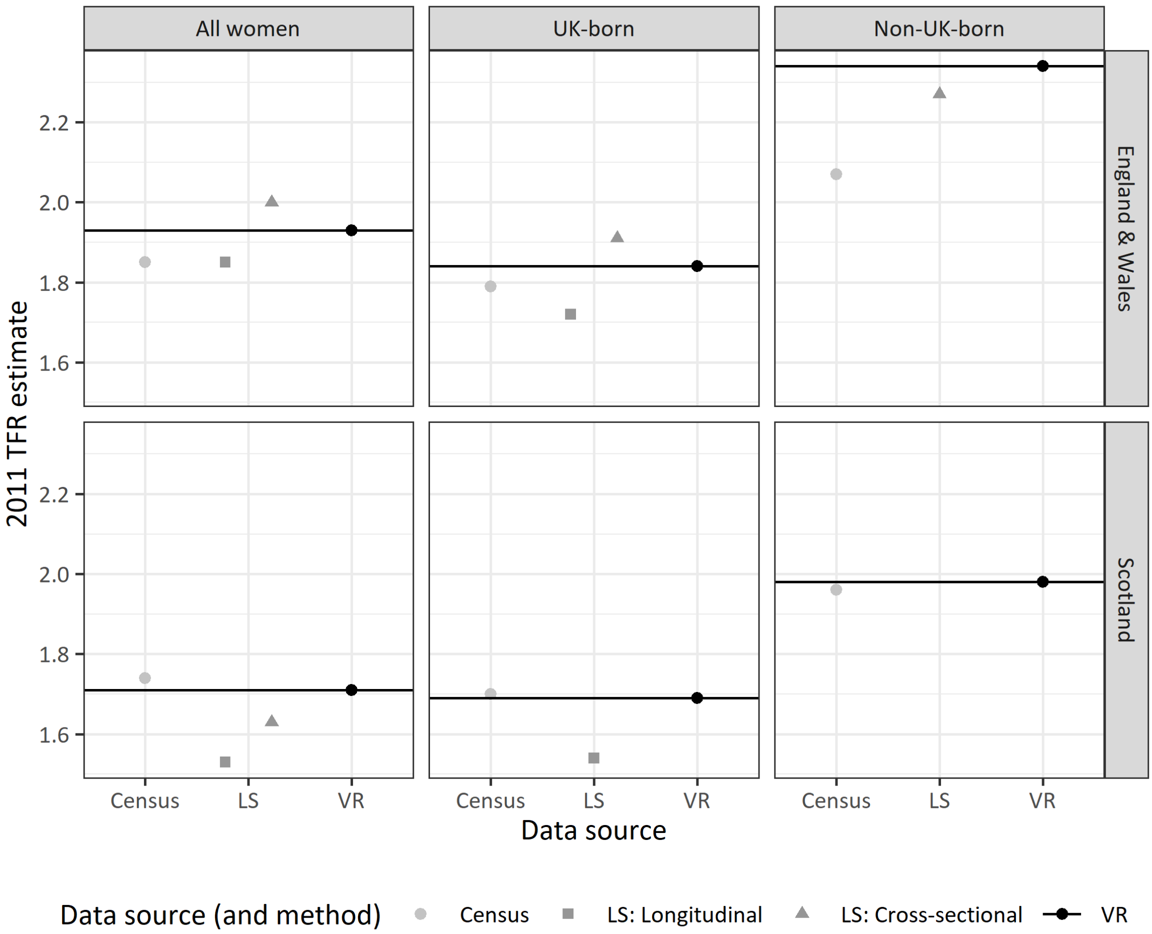

Figure 1: Estimates of the 2011 TFR for England & Wales and Scotland for all, UK-born and non-UK-born women, computed from census household microdata, longitudinal studies, and vital registration data.
Notes: For each panel, its row indicates the country (England & Wales/Scotland), and its column indicates the subsample of women (all/UK-born/non-UK-born), for which the estimates are computed. The colour indicates the data source and the shape indicates the method (LS = longitudinal study; VR = vital registration). Note that the LS estimates are computed under two methods (longitudinal and cross-sectional), while the estimates from the other data sources are computed under a single method.
Sources: NRS (2022), ONS (2019b, 2021a, 2022a, 2023a), ONS Longitudinal Study (ONS 2019c), Scottish Longitudinal Study (SLS-DSU 2023), 2011 Census Secure Household Microdata for England & Wales (ONS 2016b) and Scotland (NRS 2016).

6.5% and 3.8% for UK-born women). However, for non-UK-born women the cross-sectional estimate lies below the corresponding VR estimate (3.0% lower). For Scotland, the cross-sectional estimate is also lower than the VR estimate (4.6% lower for all women), and the longitudinal estimates show considerably larger relative decreases (10.5% and 8.9% for all and UK-born women respectively). Lastly, we note that for three of the five country-subsample combinations with census microdata and LS estimates available, the census microdata estimate is closer to the VR estimate compared to either of the LS estimates.

To explore the discrepancies between the TFR estimates from the census microdata and VR, we calculated the ratios of the numerators and denominators of the ASFRs from the two sources, for all women[6]. We found that the most extreme discrepancies were low numerator ratios, i.e. low numbers of births in the census microdata, at the younger ages. More precisely, this was up to the 30-34 age group for England & Wales, and the 20-24 age group for Scotland. We also observed that the England & Wales census ASFR estimates were noticeably lower than the VR estimates across this age range, while the Scotland estimates were very close. Taking this together with the fact that a significantly larger proportion of babies were not linked to mothers for England & Wales (4.6% vs 2.8% for Scotland) causes us to conclude that the underestimation of births at younger ages drives the greater discrepancy for England & Wales.

4. DISCUSSION

This research computes and compares estimates of the 2011 total fertility rate (TFR) for England & Wales and Scotland for all, UK-born and non-UK-born women, from three alternative data sources. We find some variation in these estimates across the data sources, with the magnitude and direction of the differences changing considerably across subsamples and countries. As both the methods used and the particular modes of data collection vary widely across sources, it is challenging to propose a simple explanation for our findings. However, we provide some discussion here, further to that given in the rest of the paper.

Regarding the vital registration (VR) estimates, although the overall TFR estimate is classified as official, the TFR estimates for UK-born and non-UK-born women are not. This is because the denominators are estimated using the Annual Population Survey (APS), which only includes private households, compared to the mid-year population estimates used for the

[6] We only consider all women because the VR numerators and denominators for the UK-born and non-UK-born subsamples are not available.

overall TFR, which also include communal establishments (ONS 2022a). The estimation of the mid-year population also has a non-negligible amount of uncertainty associated with it due to the estimation of migration, with internal migration estimated from several data sources and international migration estimated using the International Passenger Survey (IPS) (ONS 2021b). This reiterates the point made in Section 1 that there is no 'gold standard', with the VR-based TFR estimates having considerable associated uncertainty, particularly for the population subsamples.

Next we discuss the longitudinal study (LS) and census microdata estimates, both of which are based on samples of the population. Although these samples are relatively large (see Sections 2.2 and 2.3), the estimates are still likely to exhibit substantial sampling error. This is in contrast to the VR estimates, where the error is in the approximation of the population denominators. In terms of the LS's, it is of interest to contrast the cross-sectional method with the VR approach, given their similarities (see Section 2.2.3). For a random sample of the female population, all births taking place in the year of interest are included regardless of the mother's census presence, which is akin to the VR method. However, the denominator is the number of these women who are present at the time of the census, which is likely to underestimate the population at the mid-year due to non-response (and the very slight increase in population size that would have occurred in the intervening period). We would therefore expect the TFR estimates resulting from the cross-sectional method to be higher than the VR estimates, which we do indeed see for all and UK-born women in England & Wales, but not for non-UK-born women in England & Wales nor for all women in Scotland (Figure 1).

In contrast, the longitudinal method links the numerator and denominator by requiring that everyone for whom a birth would be included in the numerator is also in the denominator, and everyone in the denominator is at risk of a birth that would be included in the numerator. Therefore, in theory, this method should provide a more accurate estimate. However, this depends on the representativeness of the sample, and how this is affected by excluding women with missing or low-quality data. As the longitudinal estimates are all lower than the corresponding VR estimates (Figure 1), this suggests that the LS samples used for England & Wales and Scotland are more select than the usually resident population. It is probable that more mobile members of the population, such as migrants, for whom census or trace information is more likely to be deficient, have a higher chance of being excluded from the analysis. As their fertility is typically higher than the UK-born population, this could then lead to TFR estimates that are biased downwards. In this way it is difficult to choose between the cross-sectional and longitudinal methods, as they either tend to overestimate or underestimate the TFR relative to the VR estimate. However, the fact that the longitudinal method adheres to the principle of correspondence and can be applied to non-census years (Section 2.2.2) makes it a more methodologically sound approach and therefore marginally preferable to the cross-sectional method.

The census microdata provides a similar context to the longitudinal method, in that its accuracy depends on how representative the census household sample is of the general population, and also on whether the ability to link babies with mothers is random or exhibits some kind of pattern. In Section 3 we identified that lower-than-expected births to women aged under 35 were the key driver of the weaker correspondence of the census microdata TFR estimates with those from VR for England & Wales compared to Scotland. The fact that this issue of lower linkage rates is concentrated among the younger ages causes us to consider whether it is in some way linked to migration, and the greater proportion of births to non-UK-born women in England & Wales compared to Scotland (see Section 1). There is also the fact that even *before* removing babies from the sample due to linkage difficulties, the number of babies in the 10% sample of households as a proportion of the VR estimate of births is 9.9% for England & Wales but 10.3% for Scotland. This suggests that as well as the linkage problems, babies are slightly less likely to be picked up in the 2011 household microdata sample for England & Wales compared to Scotland. This latter point is perhaps more likely to be explained by the higher proportion of births to migrants in England & Wales and the reduced census coverage of migrants, which was also relevant for our LS discussion.

Overall, this study highlights the importance of comparing estimates of fertility measures across a range of data sources in order to get a handle on the associated uncertainty. In terms of future work, it would be of interest to extend this analysis to include Northern Ireland, which also has a longitudinal study (the Northern Ireland Longitudinal Study, NILS) that includes 28% of the population (NILS Research Support Unit 2023). This is considerably larger than the proportions included in the longitudinal studies of the other UK countries (see Section 2.2). Given the usefulness of census household microdata for estimating fertility, further analyses could also consider other census years in addition to 2011, to see if these conclusions hold across time. Lastly, the fact that the census-linked data has proved reasonably consistent with VR, and given the ability to obtain methodologically sound estimates as well as the availability of other census variables such as education and ethnicity, the longitudinal studies are an important data source for fertility analyses.

5. REFERENCES

Campisi, N., Kulu, H., Mikolai, J., Klüsener, S. and Myrskylä, M. (2020). Spatial variation in fertility across Europe: Patterns and determinants. *Population, Space and Place*, 26(4), e2308. https://doi.org/10.1002/psp.2308

Ellison, J., Dodd, E. and Forster, J.J. (2020). Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):829-856. http://dx.doi.org/10.1111/rssa.12566

Hellstrand, J., Nisén, J. and Myrskylä, M. (2020). All-time low period fertility in Finland: Demographic drivers, tempo effects, and cohort implications. *Population Studies*, 74(3):315-329. https://doi.org/10.1080/00324728.2020.1750677

Hinde, A. (2014). *Demographic Methods*. Routledge.

Jasilioniene, A., Jdanov, D.A., Sobotka, T., Andreev, E.M., Zeman, K., Shkolnikov, V.M., Goldstein, J.R., Philipov, D. and Rodriguez, G. (2015). Methods Protocol for the Human Fertility Database. *Rostock: Max Planck Institute for Demographic Research*. https://www.humanfertility.org/File/GetDocumentFree/Docs/methods.pdf

NILS Research Support Unit (2023). *About the NILS-RSU Project*. https://www.nils-rsu.co.uk/about-us/

NRS (2012). *Vital Events Reference Tables 2011 – Section 3: Births*. https://webarchive.nrscotland.gov.uk/20210314055307mp_/https://www.nrscotland.gov.uk/files/statistics/ve-reftables-2011/ve-ref-tables-2011-section3.xls

NRS (2016). *Census 2011 Household - Scotland* [Data set]. ONS SRS Metadata Catalogue, released 9 March 2016. https://doi.org/10.57906/vs5v-hw04

NRS (2022). *Vital Events Reference Tables 2021 – Section 3: Births*. https://www.nrscotland.gov.uk/files//statistics/vital-events-ref-tables/2021/vital-events-21-ref-tabs-3.xlsx

NRS (2023a). *Births – Background Information: Fertility rates*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/births/background-information

NRS (2023b). *Vital Events – General Background Information: Date of registration is basis of Vital Events statistics*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-background-information

NRS (2023c). *Vital Events – General Background Information: Geographical basis of Vital Events statistics*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-background-information

NRS (2023d). *Vital Events Reference Tables 2021: Notes and Definitions*. https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-publications/vital-events-reference-tables/2021

ONS (2012). *Births by parents' country of birth, England and Wales: 2011*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/parentscountryofbirthenglandandwales/2012-08-30

ONS (2016a). *2011 Census: Secure Microdata Files – User Guide*. https://www.ons.gov.uk/file?uri=/census/2011census/2011censusdata/censusmicrodata/securemicrodata/secureuserguidev11edcw_tcm77-381593.pdf

ONS (2016b). *Census 2011 Household - England and Wales* [Data set]. ONS SRS Metadata Catalogue, released 12 July 2016. https://doi.org/10.57906/f56d-r183

ONS (2017). *Population of the UK by country of birth and nationality: individual country data – January to December 2011*. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/populationoftheunitedkingdombycountryofbirthandnationalityunderlyingdatasheets/januarytodecember2011/underlyingdatasheetsforpopulationbycountryofbirthandnationalityjan11todec11.xls

ONS (2019a). *Births to sample mothers 1971 to 2016*. https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/paidservices/longitudinalstudyls/birthstosamplemothers1971to2016.xls

ONS (2019b). *Estimated age-specific and total fertility rates for UK-born and non-UK-born women living in households in Scotland and in Northern Ireland, 2004 to 2018*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/adhocs/10884estimatedagespecificandtotalfertilityratesforukbornandnonukbornwomenlivinginhouseholdsinscotlandandinnorthernireland2004to2018

ONS (2019c). *ONS Longitudinal Study - England and Wales* [Data set]. ONS SRS Metadata Catalogue, released 11 June 2019. https://doi.org/10.57906/z9xn-ng05

ONS (2021a). *Population estimates for the UK and constituent countries by sex and age; historical time series*. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland/mid2001tomid2020detailedtimeseries/ukpopulationestimates18382020.xlsx

ONS (2021b). *Population estimates for the UK, mid-2020: methods guide*. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/populationestimatesfortheukmid2020methodsguide

ONS (2022a). *Live births in England and Wales by parents' country of birth, 2021*. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/parentscountryofbirth/2021/workbookparentscob20212.xlsx

ONS (2022b). *Rebasing and reconciliation of mid-year population estimates following Census 2021, England and Wales: 2022*. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/rebasingandreconciliationofmidyearpopulationestimatesfollowingcensus2021englandandwales/2022

ONS (2023a). *Births in England and Wales: summary tables, 2021.* https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthsummarytables/2021/workbooksummarytables2021final.xlsx

ONS (2023b). *ONS Longitudinal Study*. https://www.ons.gov.uk/aboutus/whatwedo/paidservices/longitudinalstudyls

ONS (2023c). *User guide to birth statistics*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/methodologies/userguidetobirthstatistics

ONS (2023d). *Vital statistics in the UK: births, deaths and marriages, 2021.* https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/vitalstatisticspopulationandhealthreferencetables/current/annualreferencetables2021.xlsx

Raftery, A.E., Li, N., Ševčíková, H., Gerland, P. and Heilig, G.K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109(35):13915-13921. https://doi.org/10.1073/pnas.1211452109

Robards, J., Berrington, A. and Hinde, A. (2011). Estimating fertility rates using the ONS Longitudinal Study – what difference does the inclusion of non-continually resident members make? *Population Trends*, 144 (Summer 2011): 33-47. https://webarchive.nationalarchives.gov.uk/ukgwa/20160111181431mp_/http://www.ons.gov.uk/ons/rel/population-trends-rd/population-trends/no-144--summer-2011/estimating-fertility-rates-using-the-ons-longitudinal-study.pdf

SLS-DSU (2023). *Introduction to the Scottish Longitudinal Study*. https://sls.lscs.ac.uk/about/

United Nations Department of Economic and Social Affairs, Division for Sustainable Development Goals (2007). *Indicators of Sustainable Development: Guidelines and Methodologies, 3rd Edition - Full set of methodology sheets.* https://sdgs.un.org/sites/default/files/publications/methodology_sheets.pdf

United Nations Department of Economic and Social Affairs, Population Division (2022). *World Population Prospects 2022: Methodology of the United Nations population estimates and projections*. UN DESA/POP/2022/TR/NO. 4. https://population.un.org/wpp/Publications/Files/WPP2022_Methodology.pdf

WHO (2023). *Indicator Metadata Registry List: Total fertility rate (per woman)*. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/123

CPC
CENTRE FOR POPULATION CHANGE

CONNECTING GENERATIONS

ESRC Centre for Population Change
Connecting Generations

Building 58
Faculty of Social Sciences
University of Southampton
SO17 1BJ

Tel: +44(0)2380592579
Email: cpc@southampton.ac.uk

To subscribe to the CPC newsletter and keep up-to-date with research activity, news and events, please register online:
www.cpc.ac.uk/news/newsletter

For our latest research updates you can also follow CPC and Connecting Generations on social media:

Twitter: @CPCpopulation @ConnectingGens
Facebook: /CPCpopulation
LinkedIn: /company/esrc-centre-for-population-change-and-connecting-generations

University of Southampton · University of St Andrews · University of Stirling

Leverhulme Centre for Demographic Science · Resolution Foundation · Office for National Statistics · National Records of Scotland

Producing novel science
Informing policy
Improving knowledge on population change

cpc.ac.uk

ISSN 2042-4116