Sex ratios and Residual Bias adjustment for 2021 Census

Key Message

While the last two Censuses have had robust estimation processes the sex ratios produced have appeared implausible. The issue of sex ratio patterns identified in the last two censuses is likely to be an issue in the 2021 Census as the estimation approach is broadly similar.

The adjustment made in 2011 has performed quite well at a national level over the decade, but there is still evidence in the mid-year population estimates of simply ageing forward the sex ratio found in the census, which suggests that our estimation of international migration by age and sex is likely not correct, or the sex ratio derived from census is still not correct, or possibly both.

Recommendation

The minimum approach to providing an adjustment is the use of the ONS Longitudinal Study (LS) to provide an alternative set of age specific sex ratios and the ONS LS is set up to do this. The ONS LS contains linked Census and life events data for a 1 per cent sample of the population of England and Wales.[1] Judgement can then be used to adjust the 2021 Census estimation if the result produced by initial estimation is considered sub-optimal.

Discussion at the estimation and adjustment working group in February 2021 suggested that any distribution of such an adjustment should be done from within the Census estimation process rather than calibrated to an external source. A separate paper methodology developed for distribution will be produced.

Work will continue on exploring what may be possible with administrative data, both as a national level comparator and as a potential way of providing a sensible comparator for the distribution for any residual bias adjustment.

Feedback needed from CRAG

Agreement that the ONS LS model is used to assess whether a residual bias adjustment is required for the 2021 Census.

Agreement that administrative data is not used for producing any national adjustment nor distributing subnationally any adjustment required.

Further work needed

Presentation of the method of distributing any national adjustment to local authority level.

Further collation of data over the decade to provide evidence of sex ratios, and changes in sex ratios over time, and sex age balance of international migration over time.

[1] ONS Longitudinal Study

Introduction

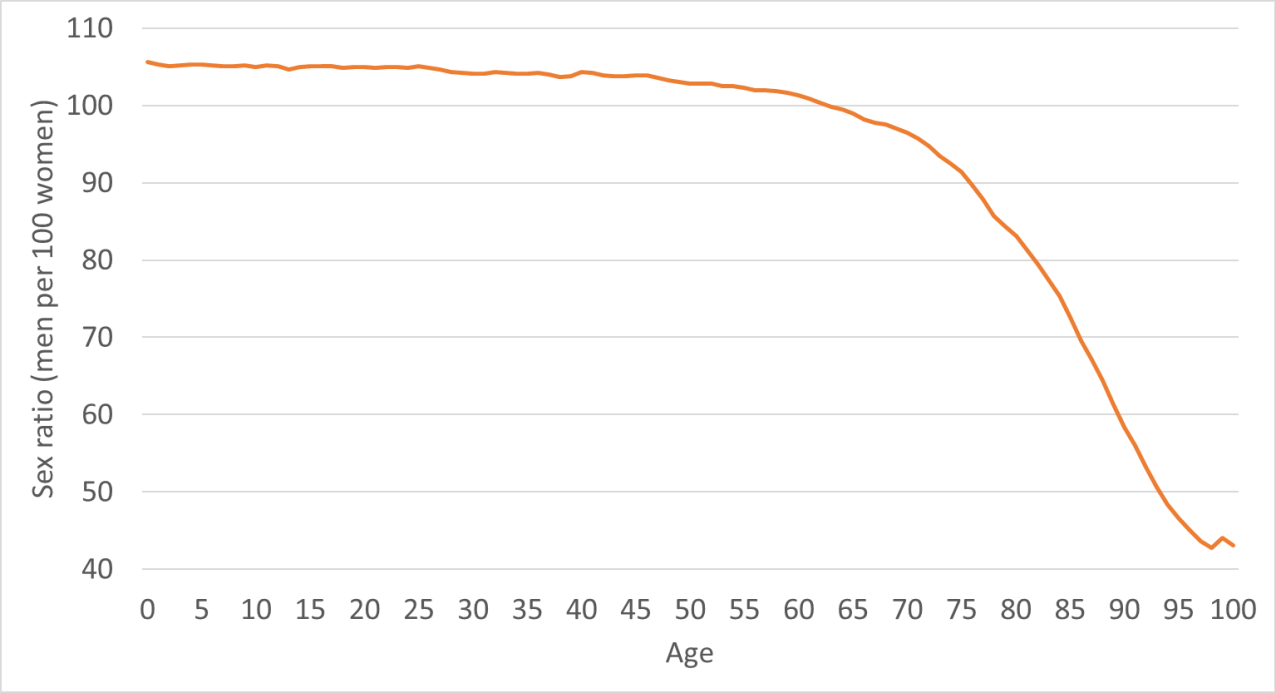
For both the 2001 and 2011 Censuses sex ratios have played an important role in understanding their quality. After the 2001 Census an adjustment was made to the mid-year estimates based on the census and in 2011 an adjustment was made to the census estimation process.

This paper provides an update on the use of sex ratios to make any residual bias adjustment that may be required for the 2021 Census. The paper discusses the LS adjustment providing:

- a short section describing the methodology and theory of the model
- an update on the replication of the model so that it is ready for use in 2021
- some sensitivity testing around changing the assumptions in the model.

Unusual patterns in sex ratios are mainly related to patterns of migration and/or the capture and estimation of young adults in census. Without any migration, an estimated sex ratio is easy to derive from births and mortality rate data over the past 100 years[2] – see figure 1. This shows a sex ratio that pretty much reflects the sex ratio at birth of around 105 males per 100 females until there is sufficient differences in levels of mortality by sex ages at above 50 to start decreasing the sex ratio.

Figure *1* Sex ratio derived from births and mortality rates in the last 100 years, England and Wales, 2019



Source: ONS Vital events data and mortality database, authors own calculation

[2] Steve Smallwood and Sofie De Broe (2009) Sex ratio patterns in population estimates Population Trends no 137 https://link.springer.com/content/pdf/10.1057/pt.2009.33.pdf

The patterns found in the Census, and in administrative data, are very different from the natural sex ratio in figure 1. Differences can only be a result of differential net migration by males and females or from differential issues with estimation of males and females in the census. Therefore, there are only two immediate solutions to demographically assessing the sex ratio. Examining as many sources as possible of international migration to look for sufficient differences by age and sex in net migration flows to support the changing sex ratio and find an alternative way of assessing the non-response in Census by age and sex at the national level to produce a more robust sex ratio pattern.

In this paper some evidence is presented on data being collated on migration, but this work is as yet incomplete and will be added to.

Finding an alternative assessment of non-response and therefore an alternative national sex ratio has been carried out in the last two census rounds by using the ONS Longitudinal Study (ONS LS). The decision at this stage is about continuing to use the ONS LS as an approach. The decision on whether and by how much an adjustment is made will be based on the empirical data from the 2021 Census, including estimation made and the ONS LS link. A further complication is that within the last two years we have had two seismic events Brexit and the Pandemic, which may have impacted in unusual ways on migration patterns.

As an office we are still working to understand the impact on the levels of migration, and we have not yet even considered the detail around age and sex. The Pandemic has also had a major impact on living arrangements in England and Wales. This may impact on coverage and response in the 2021 Census too. The phrase 'new normal' may be overused, but much of demography is based on the usual – we do not yet understand how all that has happened impacts on that usual and what the new usual may be. These events will also have undoubtedly affected the way that administrative data behaves because of changes in operations for those operating administrative systems; changes in the way people interact with admin data; and, actual societal changes resulting from the pandemic. Elements of the ONS LS are based on administrative data from the registration system and from the NHS.

Section 1 of this paper (and Annex A) provides some information on the ONS LS attrition model. Section 2 provides some contemporary international comparisons. Section 3 begins to show some evidence around collated intercensal sources of migration. This section is currently incomplete and will be added to in the next few months. Finally, section 4 of the paper briefly examines sex ratio patterns from other data sources. This confirms that these do not provide a useful source for consideration for adjustment at this stage.

1. Updated evidence around the ONS Longitudinal Study attrition model

The ONS Longitudinal Study (LS) was used to help derive an adjustment for the 2011 Census.[3] It was also used to derive an adjustment for the 2001 mid-year estimates because the estimation of males in the 2001 Census was considered deficient.

The essence of the adjustments derived is that the LS provides an independent check on response as it can be used to consider the number of people you expect to be part of the study and compare it with the number of people subsequently found in the latest census link.

However, the Longitudinal Study, being a linkage study, might expect some attrition from linkage error and other issues, as well as census non-response. When the model was set up it was considered that the best measure of pure attrition was for the period 1971 to 1981. This was because the response rate to the Census was considered to be very high, over 99%.[4] Attrition rates calculated from that year are then applied in a chained approach to produce an expected number of responses for the latest Census to be linked.

This gives an expected number of LS members by 5-year age group and sex. These figures can then be compared with the actual number found. This gives an alternative implied response rate by age and sex.

The Attrition rates used in the model are calculated for three separate groups.

(a) Those that are found in the previous census (adjusted for deaths and emigration),
(b) Those that are born in the intercensal period (adjusted for deaths and emigration), and
(c) Those that immigrate into the LS in the intercensal period (adjusted for deaths and emigration)

The rates are illustrated in Figure 2, alongside the rates derived for subsequent intercensal periods using the model. This shows how overall attrition in subsequent years is most likely affected by greater levels of non-response than the 1971 to 1981 period (which is assumed to be least affected by non-response) and also the changing composition of the ONS LS with increased migration.

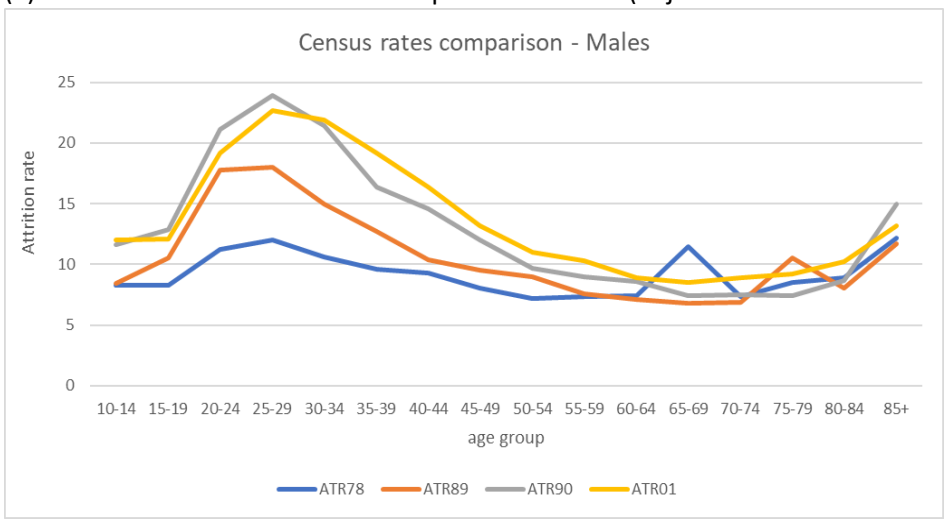
This can then be used to produce an alternative population estimate from the Census count. In 2001 the difference seen for males which occurred mainly at ages 25-34 was used to generate an adjustment to the mid-year estimates. In 2011 the overall sex ratio produced by these alternative estimates was used to recalibrate the 2011 Census adjustment.

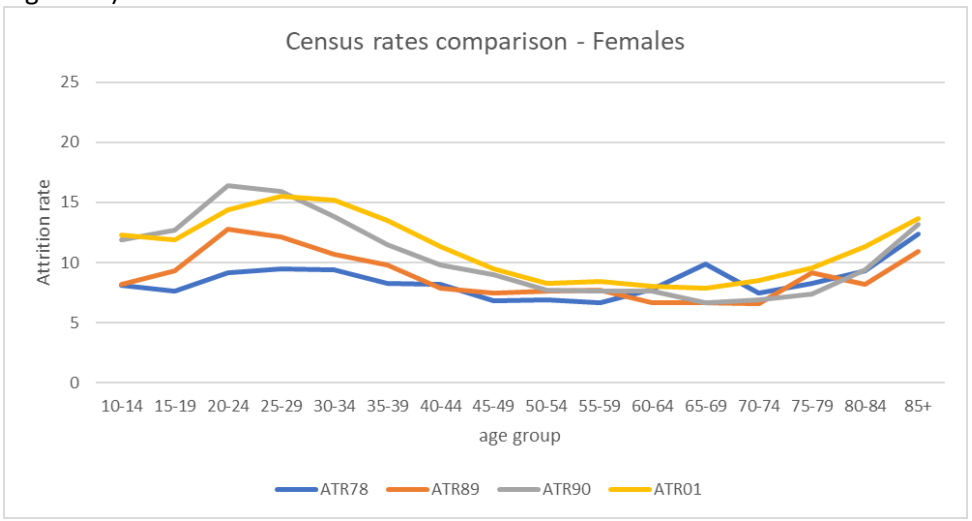
A detailed description of how the model is built is found in Annex A.

[3] Office for National Statistics (2012k) Making a national adjustment to the 2011 Census.
[4] From History, Organisation and Quality of data ONS LS 1971-1991. Chapter on Census quality

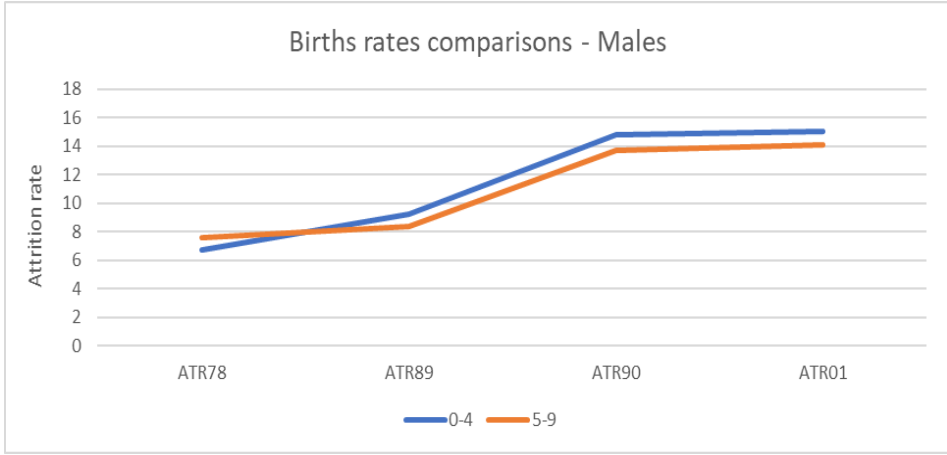
Figure 2 Intercensal attrition rates 1971 to 1981 and modelled attrition rates for subsequent intercensal periods based on 1971-1981 attrition

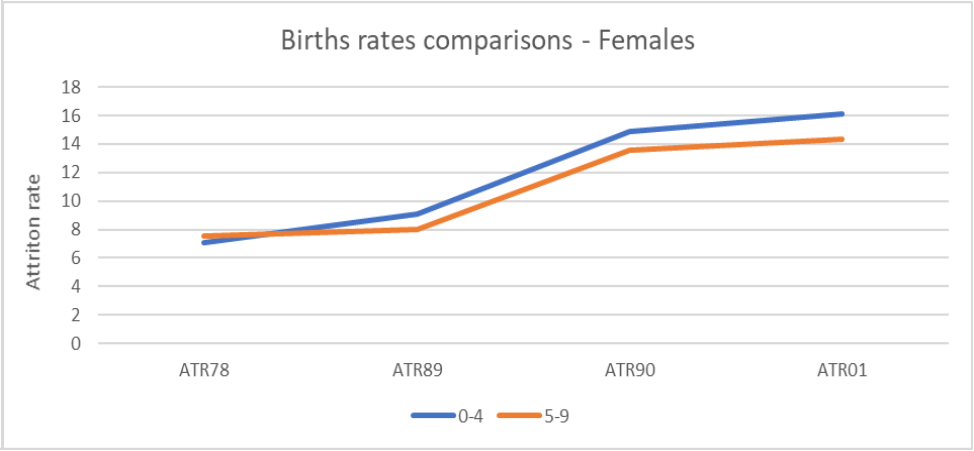
(a) Those that are found in the previous census (adjusted for deaths and emigration)



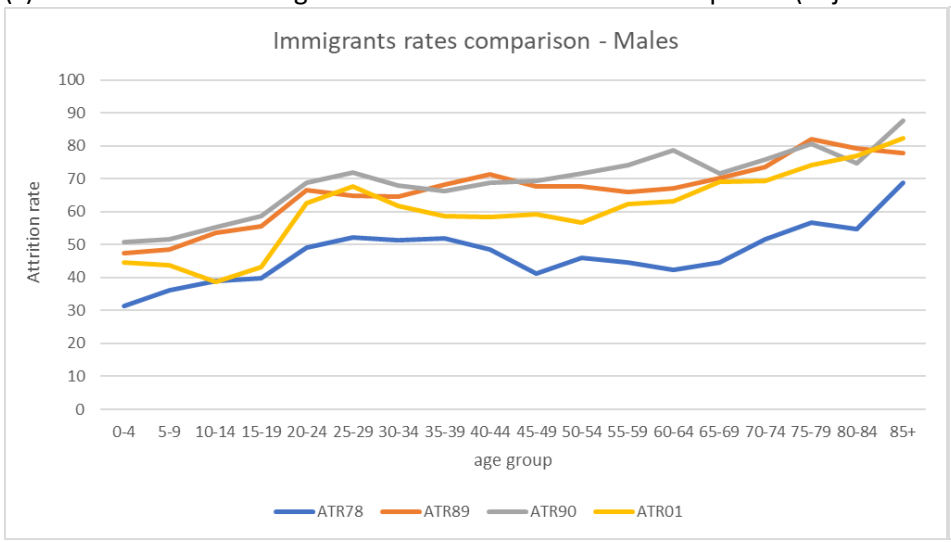


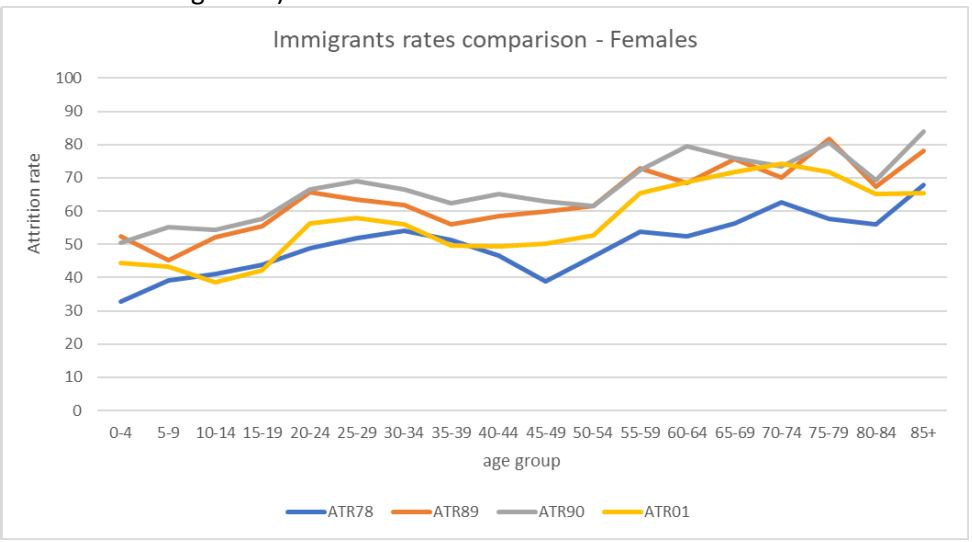
(b) Those that are born in the intercensal period (adjusted for deaths and emigration)





(c) Those that immigrate into the LS in the intercensal period (adjusted for deaths and emigration)





Source for all Figure 2 charts: ONS Longitudinal Study

It should be noted that when the model is implemented calculated attrition proportions become similar to the actual derived intercensal attrition for the individual intercensal decades.

The ONS LS team have been ensuing they have the code and ability to run the model in 2021. The team have replicated the initial 1971 to 1981 attrition rates and the subsequent analysis to 2011 is almost the same as produced at around the time of the 2011 Census. As the dataset is live there are very small changes. They will be ready to operate the model in 2021. The differences in the derived sex ratios are shown in Table 1. A reminder that these sex ratios were used with a simple weighting approach in the 2011 adjustment – giving full to the sex ratio for age 20-24, 5/6ths to 25-29 and so on up to age group 45-49.

Table 1 – a comparison of the outcome target sex ratios (Men per 100 women)with rerun on the ONS Longitudinal Study

Age Group	Sex ratio (LS) 2011	Sex ratio (LS) 2011 Rerun	Difference in the Sex Ratios
0-4	104.043	103.976	0.067
5-9	105.004	104.975	0.029
10-14	104.577	104.539	0.038
15-19	104.154	104.092	0.062
20-24	101.795	101.765	0.030
25-29	100.917	100.892	0.026
30-34	102.493	102.511	-0.018
35-39	101.153	101.200	-0.047
40-44	100.066	100.030	0.036
45-49	99.408	99.333	0.075
50-54	99.985	99.936	0.049
55-59	98.211	98.172	0.040
60-64	97.227	97.240	-0.013
65-69	93.117	93.123	-0.005
70-74	90.654	90.602	0.052
75-79	82.546	82.507	0.039
80-84	58.833	58.773	0.060

Source: ONS Longitudinal Study

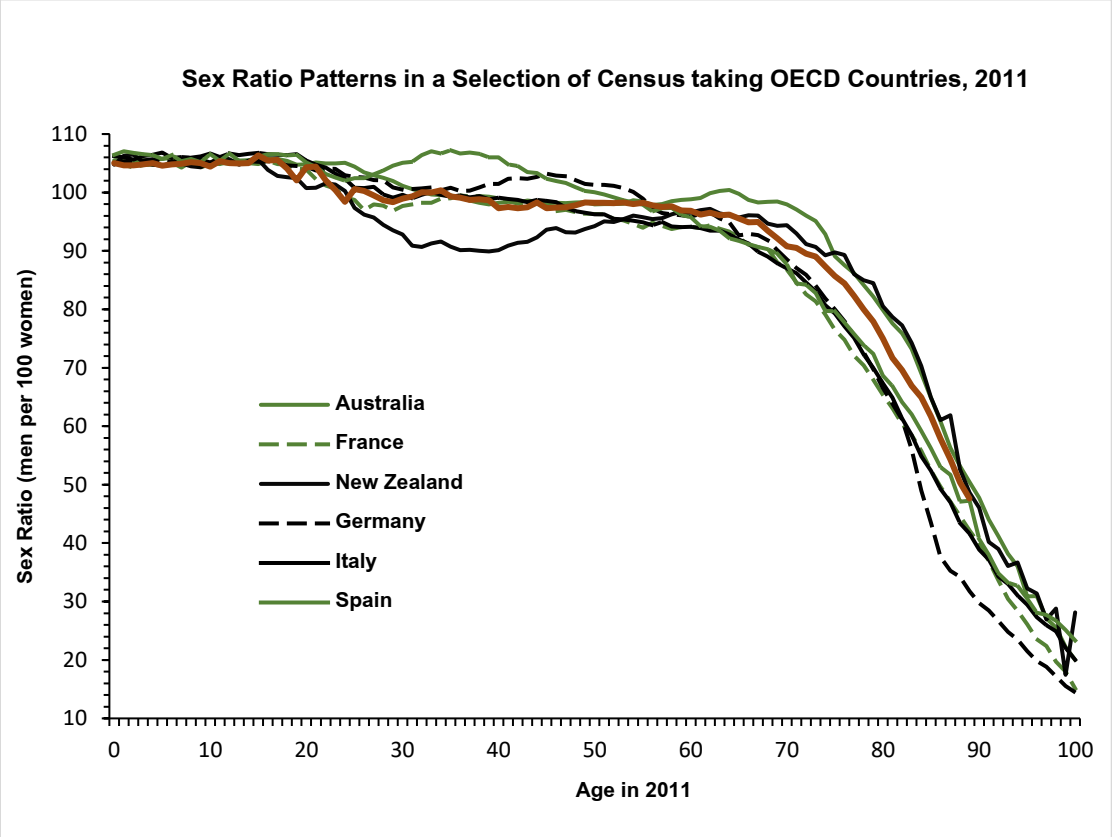
Delays in processing mean that there are some issues about having all the data for the 2011-2021 decade – particularly around births and deaths where it is likely that not all events will be available up to census day. This is disappointing but not critical as births are not relevant to the calculation of adult sex ratios, where sex ratios are most likely to be of concern, and deaths mostly occur at ages above the age groups we are interested in. Further, some events (deaths, embarkations, and re-entries) that will not be available from processed LS data will be available from NHS events data up to 2021 Census Day.

2 International Comparisons

Previous analysis for data from around the 2001 Census period prior to the 2011 Census showed that there was more variation in sex ratios in Census taking countries than those that operate a population register.[5] Also, there was no evidence of a rapid change in sex ratios around ages 18-20. The following charts show sex ratios for selected countries around the year 2011, when the previous round of Censuses was held. The findings from De Broe and Smallwood still hold.

Note in Figure 2 the sex ratio is based on the mid-year estimates for 2011, which in turn are based on the 2011 Census with the residual bias adjustment (formerly known as national) applied. Without this the ratio for England and Wales would have dipped as low as 95 at ages in the late twenties.[6]

Figure 2a

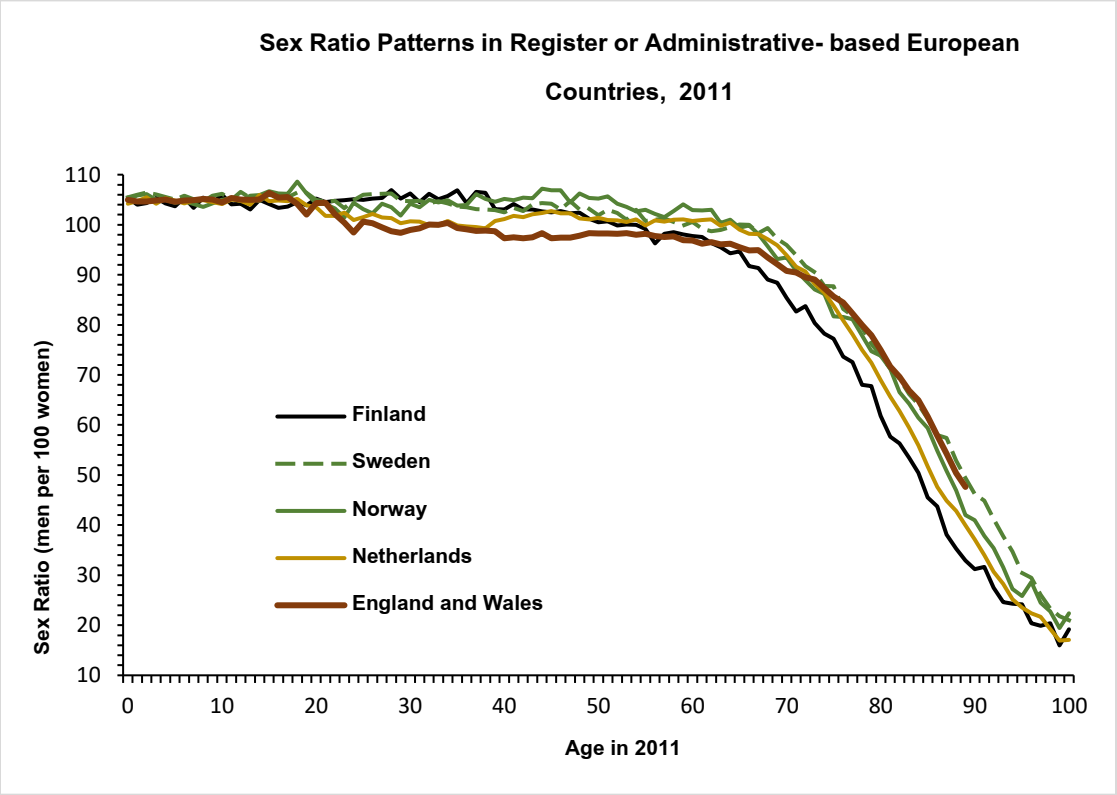


Source: OECD database

[5] Steve Smallwood and Sofie De Broe (2009) Sex ratio patterns in population estimates Population Trends no 137 https://link.springer.com/content/pdf/10.1057/pt.2009.33.pdf

[6] Office for National Statistics (2012k) Making a national adjustment to the 2011 Census. Available at: http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/making-a-national-adjustment-for-residual-biases.pdf

Figure 2b



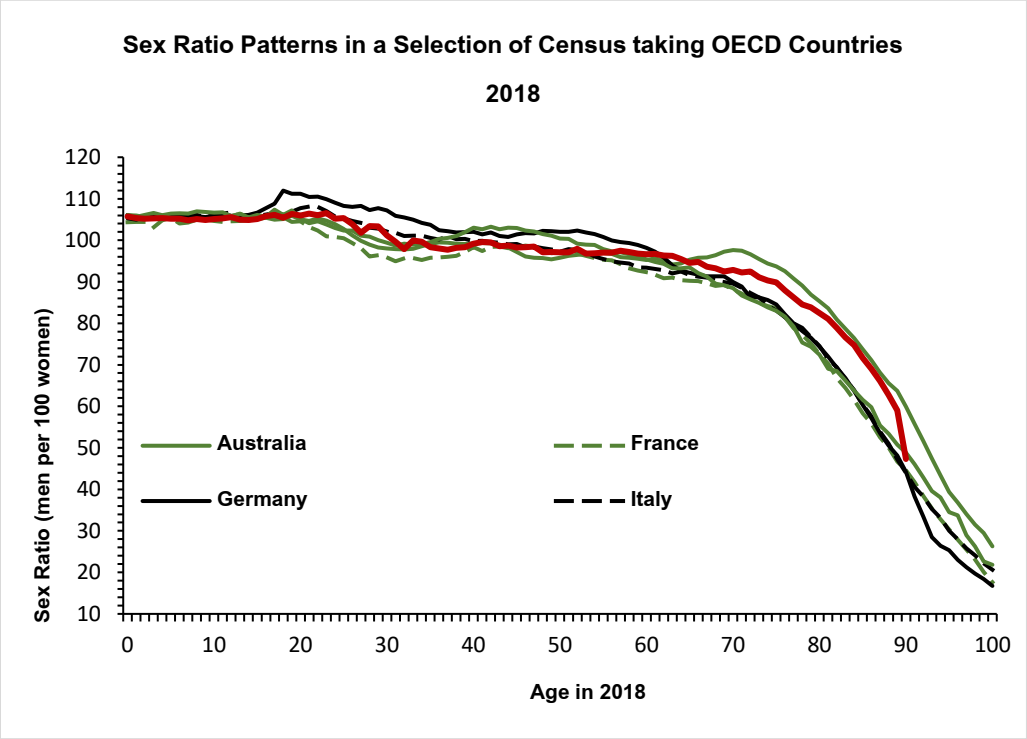
Source: OECD database

Figure 2a shows that in general census taking countries appear to have quite variable sex ratios. Figure 1b shows sex ratios from register-based countries. The England and Wales line on both graphs is the 2011 Mid-year estimates sex ratio based on the adjusted 2011 Census figures. As was previously argued in De Broe and Smallwood, this may be the result of migration, particularly emigration not being well measured. It will also reflect the types of country and society but does provide some evidence that registers have more stable and explainable sex ratios with migration insufficient to move sex ratios by much in the young adult ages.

Figures 3a and 3b show some later information, which continues to demonstrate the relative stability of the register countries. The unusual spike in sex ratios in Sweden is an effect of migration that is visible in their migration data. It is mianly driven by asylum seeker settlements, where over 90 per cent were male.[7]

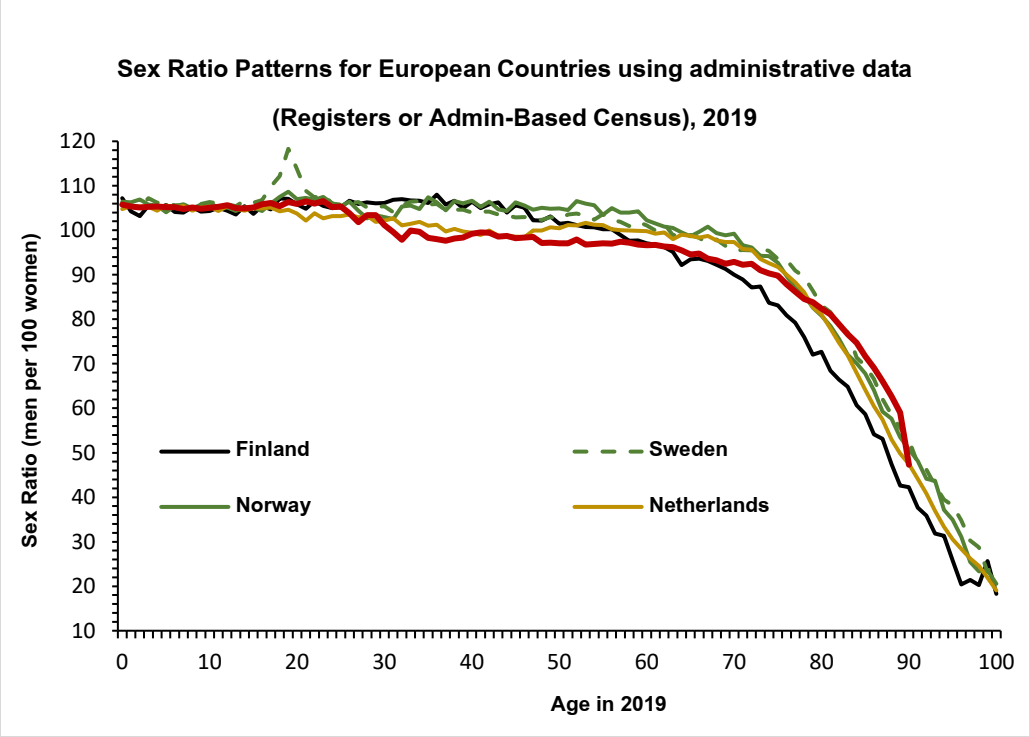
[7] Correspondence with statistics Sweden "The spike of boys/young men is due to a large number of unaccompanied minors who sought asylum in 2015. In total they were around 35 000 and more than 32 000 of them where boys, many just under the age of 18. Due to a large number of people seeking asylum in 2015 the Migration Agency were not able to process the applications in time. A special set of regulations were introduced for this group and many of them have been granted residence permits and been registered as immigrants in 2016–2019."

Figure 3a



Source: OECD database

Figure 3b



Source: OECD database

3. Information from Sources of International migration in the decade 2011 to 2021.

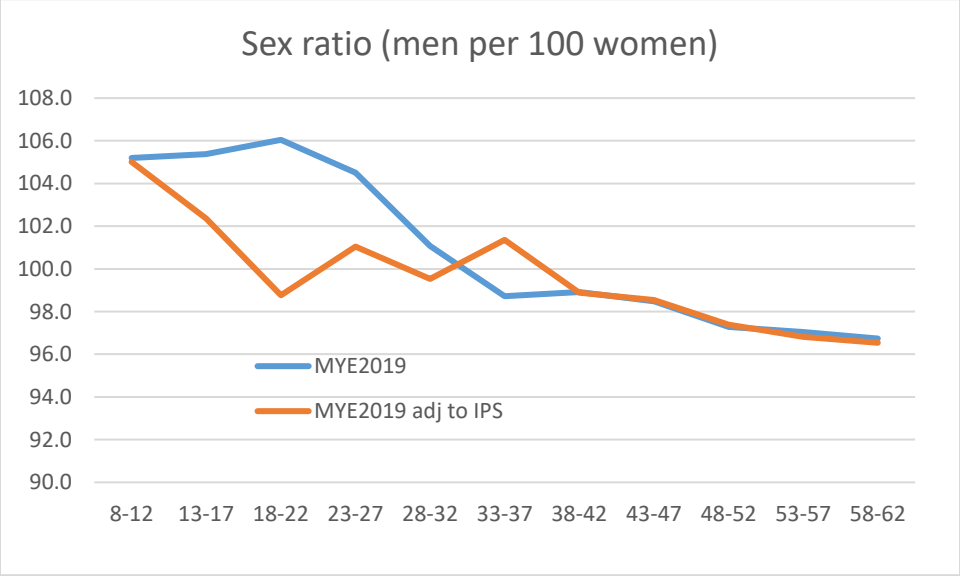
In addition to looking at sex ratio patterns other evidence used in 2011 was on migration patterns over the decade to look for difference in patterns between males and females. Table 2 collates data that has been put together so far. There are two ways of looking at cohort change from migration: any direct measure of migration by age cohort and sex, and/or looking at the change in cohorts by age and sex, with an adjustment for deaths, to imply net migration.

The table is still to be completed for several datasets, but it is possible to compare the age sex patterns used in our mid-year estimates in the decade to raw migration estimates from the International Passenger survey.

As can be seen from the table, for the ages we are interested in deaths can be regarded as too small to have much of an effect. The figures for migration used in the mid-year estimates do not use the age information from the IPS for inflows. They do use it for outflows and also the sex split in international migration is derived from the IPS. The MYE figures do not greatly change the sex ratio of the population when applied. However, if we replaced them with the age-sex distributions from the raw IPS data there is an effect (see Figure 4). Indeed, it would bring the sex ratio closer to that seen in the Census and stop the ageing forward of the drop seen around ages 18 to 20 in the 2011 Census. There should be some caution about using the raw age-sex data from the IPS as we know that a number of other biases and issues have been found with that data, however it may be indicative of a change in sex patterns of migration in the last decade. It should be noted that such imbalances are not seen in the Migrant Worker Scan data provided in the table, although this includes some short-term migration and obviously does not cover all migrants, particularly those that migrate purely as part of families or for study.

Work is ongoing to collect more information on intercensal migration flows and also to infer net migration flows from changes in stocks of administrative data.

Figure 4 Adjusted Mid-year estimate sex ratio using raw IPS age sex data rather than methods used in Mid-year estimates (age in 2019)



Source: ONS Mid-year estimates and International Passenger Survey

Table 2 summary of cohort change from various sources

Table of population change by cohort Mid-2011-2019, England and Wales																								
- Age at mid-2011																								
	0-4		5-9		10-14		15-19		20-24		25-29		30-34		35-39		40-44		45-49		50-54		55-59	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
International Migration assumed in mid-year estimates																								
(a) Inflow	108,253	103,870	89,902	81,200	169,201	147,897	409,603	381,003	464,903	406,124	304,549	241,709	187,651	146,633	126,773	100,615	94,164	73,700	71,615	57,126	53,452	45,304	44,490	40,145
(b) Outflow	21,172	24,340	14,460	16,165	34,665	41,570	211,804	200,521	340,566	268,236	258,802	166,199	164,663	107,447	103,874	68,302	63,425	41,076	45,914	28,128	34,578	22,170	27,427	15,891
IPS (MS adjusted)																								
Inflow	37,071	26,594	29,647	29,868	130,181	178,220	396,739	475,679	487,299	484,791	311,219	281,773	205,224	136,703	114,323	95,252	65,763	51,988	47,262	38,020	26,011	15,799	17,955	18,802
Outflow	20,642	22,468	13,970	15,324	19,329	31,342	186,604	167,310	317,438	266,898	232,022	158,009	151,778	105,235	99,013	63,094	58,485	36,787	31,341	22,931	33,468	16,443	21,195	14,872
Republic of Ireland																								
Inflow																								
Outflow																								
Asylum																								
Inflow																								
Total Inflow																								
Admin sources inflow:																								
MWS*							401,101	405,944	401,101	405,944	730,536	698,160	332,552	246,584	227,231	166,281	154,510	117,500	87,954	71,373	47,938	42,736	9,794	9,125
MWS inc ST																								
Flag 4*																								
Total Outflow																								
Approx Cross border net flow																								
Net flow MYE	87,081	79,530	75,442	65,035	134,536	106,327	197,799	180,482	124,337	137,888	45,747	75,510	22,988	39,186	22,899	32,313	30,739	32,624	25,701	28,998	18,874	23,134	17,063	24,254
Net Flow - IPS (MS adjusted)	16,429	4,126	15,677	14,544	110,852	146,878	210,134	308,368	169,861	217,893	79,197	123,763	53,445	31,468	15,310	32,158	7,278	15,201	15,921	15,089	- 7,457	- 644	- 3,240	3,930
Deaths	3,313	2,552	1,215	893	2,703	1,514	5,730	2,484	8,301	3,734	10,551	5,561	14,395	8,456	20,492	12,575	34,812	22,509	51,979	34,851	70,877	48,854	98,452	67,481

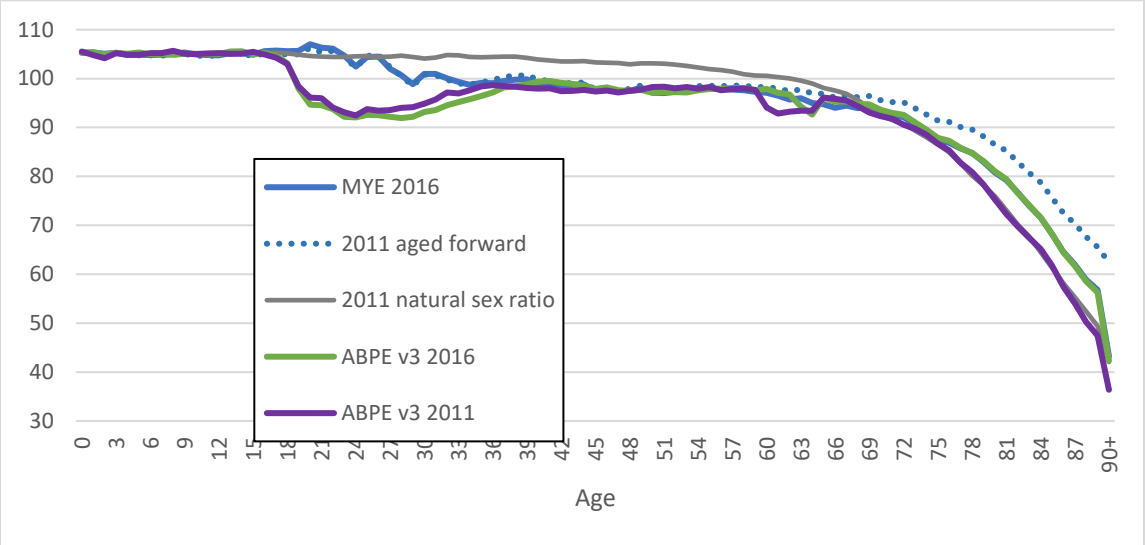
[More data to be added above as we get it]

4. Other sex ratios

In 2011 the LS was deemed to be the only reliable source to use as stock level use of admin data across the whole population was rare (change in stocks has been used for internal migration, and HESA stock information used to adjust that). One particular source, the Lifetime Labour Market (L2) database, was considered but not used. This did provide a similar sex ratio pattern to that derived from the LS.

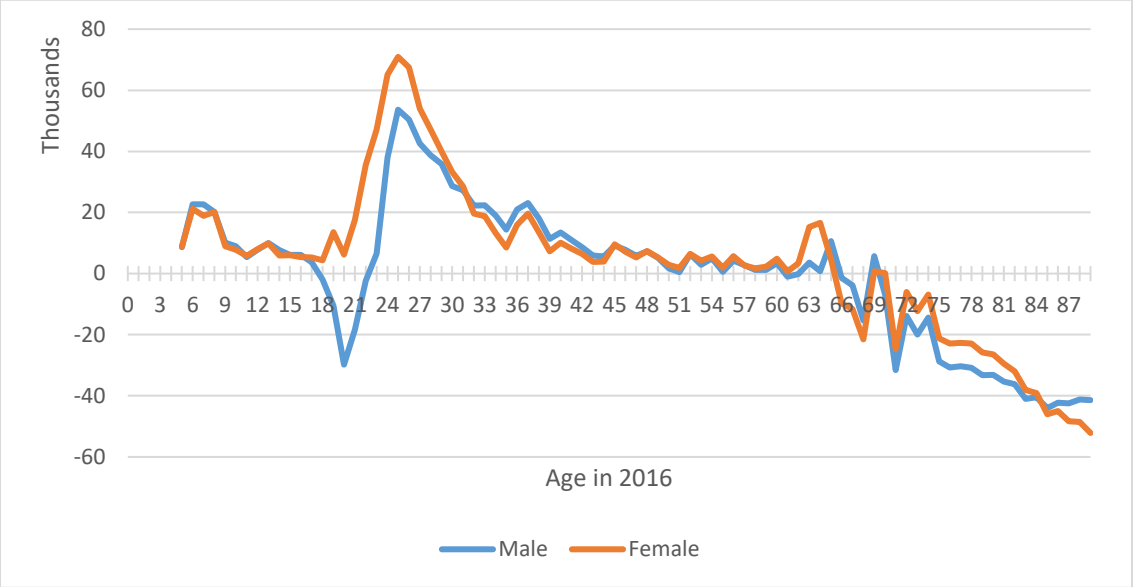
ONS has also been experimenting with producing administrative based population estimates. The latest set of these has been based on activity and is intended to have coverage adjustment applied. Data for 2019 will be available shortly. However, the published data for 2011 and 2016 shows very unusual sex ratio patterns. These were noted in paper EAP126 presented to the external panel and are reproduced below for reference. The first chart (Figure 5) shows selected sex ratio patterns. The second (Figure 6) shows the implied cohort change between 2011 and 2016. At present this does not seem robust. We will continue to monitor this work and also explore the data from DWP/HMRC where an activity-based process has been used to produce a dataset that is designed to be representative of the whole population called RAPID.[8]

Figure 5 Sex ratios from Administrative based population estimates and mid-year estimates for 2011 and 2016



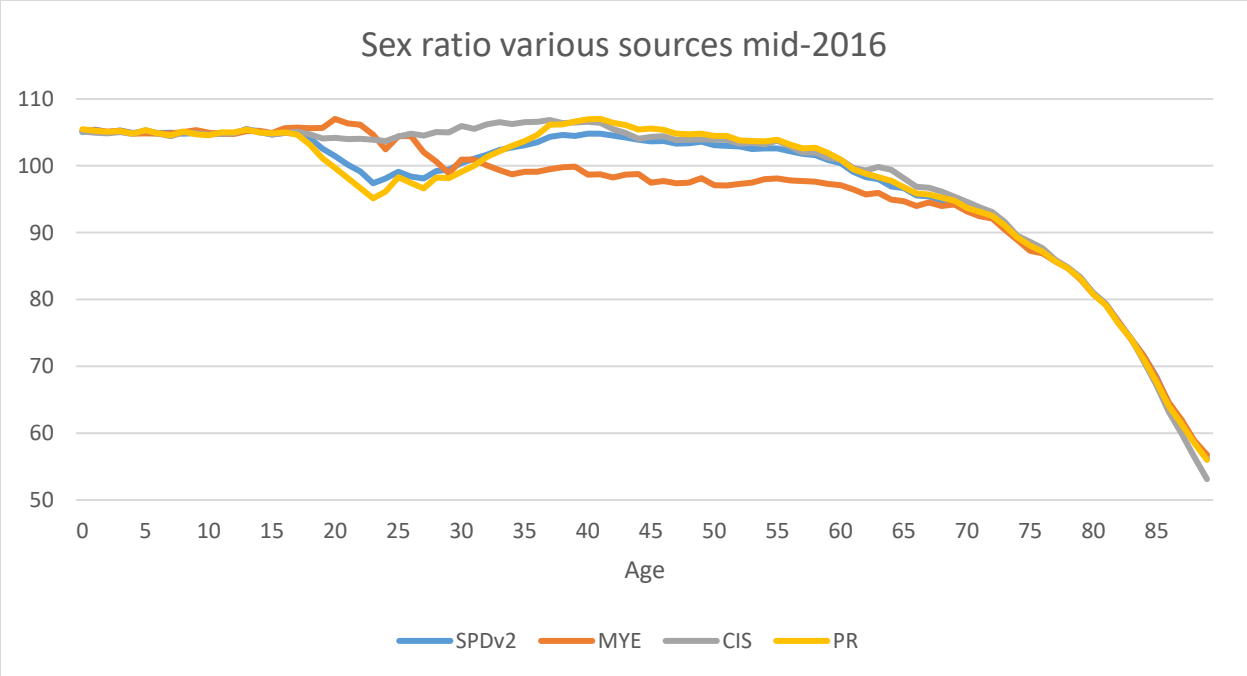
[8] See section 3 of International migration statistics preliminary adjustments update and feedback: February 2020https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/internationalmigrationstatisticspreliminaryadjustmentsupdateandfeedback/february2020

Figure 6 Implied net population change by cohort from ABPEv3 between 2011 and 2016



The sex ratio in various other sources for a selected year is shown below. The sex ratio in the ABPE is likely to be driven by the sex ratio in the patient register. Here we have compared to the earlier version of the admin data population estimates (SPDv2) which was driven by presence on datasets rather than activity. However, it tells a similar story.

Figure 7 Sex ratios from Patient Register, DWP/HMRC Customer information system, SPDv2 (now known as ABPEv2) and mid-year population estimates.



Annex A - **Modelling LS attrition at 1971/81 attrition rates**
(With grateful thanks to Kevin Lynch for the original documentation and Caroline Graham for the update to 2021)

The approach can be understood incrementally, considering attrition at different censuses:

Time period	t
1971 Census	0
1981 and preceding decade (after 1971 Census)	1
1991 and preceding decade (after 1981 Census)	2
2001 and preceding decade (after 1991 Census)	3
2011 and preceding decade (after 2001 Census)	4
2021 and preceding decade (after 2011 Census)	5

1971/81 attrition

Observed attrition in 1981 has three components:
$Att_1 = r_1 + h_1 + n_1$

Where r_1 = number of traced LS members found in 1971 but not in 1981,
h_1 = number of LS members entering as NHS Central Register (NHSCR) immigrants in the 1970s but not found in 1981
n_1 = number of LS members born between 1971 and 1981 Census (1970s) but not found in 1981.

The 71/81 census-to-census attrition rate for traced LS members found in the 1971 Census but not accounted for in 1981 can be represented as:

$$\frac{r_1}{N_0 - s_1}$$

Where N_0= number of LS members found and traced in 1971,
s_1= number of traced LS members found in 1971 but either died or embarked (exits) before 1981.

The 1971/81 immigrants attrition rate for LS members who entered as NHSCR immigrants in the 1970s but were not found in 1981 can be represented as:

$$\frac{h_1}{I_1 - k_1}$$

Where I_1 = number of LS members entering as NHSCR immigrants in the 1970s,
k_1= number of LS members entering as NHSCR immigrants in the 1970s but having died or embarked by 1981.

The 1971/81attrition rate for new births is

$$\frac{n_1}{B_1 - l_1}$$

Where B_1= number of LS members born in the 1970s
l_1= number of LS members born in the 1970s but having died or embarked by 1981.

1981/91 attrition

The 1971/81 attrition rates are used to estimate LS attrition for 1981/1991, now with four components:

$Attr_2 = r_2 + h_2 + n_2 + m_2$

r_2 = attrition by 1991 of LS members found in 1981, at 71/81 census-to-census attrition rates:

$$r_2 = \frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2\right)$$

s_2= number of observed 1980s deaths and embarkations for LS members found in 1981.

In addition, attrition by 1991 of LS members entering as NHSCR immigrants over the 1980s but not found in 1991, at 1970s immigrants attrition rates:

$$h_2 = \frac{h_1}{I_1 - k_1}\left(I_2 - k_2\right)$$

Where I_2 = number of LS members entering as NHSCR immigrants in the 1980s
k_2 = number of LS members entering as NHSCR immigrants in the 1980s but having died or embarked by 1991.

Attrition by 1991 of LS members born in the 1980s but not found in 1991, at 1970s new births attrition rates, is represented as:

$$n_2 = \frac{n_1}{B_1 - l_1}\left(B_2 - l_2\right)$$

Where B_2 = number of LS members born in the 1980s
l_2 = number of LS members born in the 1980s but having died or embarked by 1991, and
n_2 = number of LS members born in the 1980s but not found in 1991.

For 1991, there is also attrition of traced LS members found at the 1981 Census for the first time.

Let M_1 = number of LS members entering and traced in 1981.

The attrition rate for this group is assumed to be the same as for other LS members found at the 1981 Census. Then attrition by 1991 is represented as

$$m_2 = \frac{r_1}{N_0 - s_1}\left(M_1 - d_2\right)$$

Where d_2 = number of observed 1980s deaths and embarkations for LS members entering and traced in 1981.

Taken together, 1981/91 attrition is:

$$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2\right) + \frac{h_1}{I_1 - k_1}\left(I_2 - k_2\right) + \frac{n_1}{B_1 - l_1}\left(B_2 - l_2\right) + \frac{r_1}{N_0 - s_1}\left(M_1 - d_2\right)$$

1991/2001 attrition

Attrition for 1991/ 2001 also has four components:

$Attr_3 = r_3 + h_3 + n_3 + m_3$

r_3 = attrition by 2001 of LS members found in 1991, using 71/81 census-to-census attrition rates:

$r_3 =$

$$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 - m_2 - s_3\right)$$

s_3= number of observed 1990s deaths and embarkations for LS members found in 1991.

In addition, attrition by 2001 of LS members entering as NHSCR immigrants over the 1990s but not found in 2001, at 1970s immigrants attrition rates:

$$h_3 = \frac{h_1}{I_1 - k_1}\left(I_3 - k_3\right)$$

Where I_3 = number of LS members entering as NHSCR immigrants in the 1990s

k_3 = number of LS members entering as NHSCR immigrants in the 1990s but having died or embarked by 2001.

Attrition by 2001 of LS members born in the 1990s but not found in 2001, at new births attrition rates observed between 1971/81, is represented as:

$$n_3 = \frac{n_1}{B_1 - l_1}\left(B_3 - l_3\right)$$

Where B_3 = number of LS members born in the 1990s

l_3 = number of LS members born in the 1990s but having died or embarked by 2001, and

n_3 = number of LS members born in the 1990s but not found in 2001.

For 2001, there is also attrition of traced LS members found at the 1991 Census for the first time.

Let M_2 = number of LS members entering and traced in 1991

Attrition by 2001 is represented as

$$\frac{r_1}{N_0 - s_1}\left(M_2 - d_3\right)$$

Where d_3 = number of observed 1990s deaths and embarkations for LS members entering by, and traced in 1991.

Thus, 1991/01 attrition is:

$$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 - m_2 - s_3\right)$$
$$+ \frac{h_1}{I_1 - k_1}\left(I_3 - k_3\right) + \frac{n_1}{B_1 - l_1}\left(B_3 - l_3\right) + \frac{r_1}{N_0 - s_1}\left(M_2 - d_3\right)$$

2001/2011 attrition

Attrition for 2001/ 2011 also has four components:

$Attr_4 = r_4 + h_4 + n_4 + m_4$

Here, r_4 represents attrition by 2011 of LS members found in 2001, at 71/81 census-to-census attrition rates:

r_4 =

$$\frac{r_1}{N_0 - s_1}\left(\begin{array}{l} N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 \\ - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4 \end{array}\right)$$

s_4= number of observed 2000s deaths and embarkations for LS members found in 2001.

In addition, attrition by 2011 of LS members entering as immigrants over the 2000s (NHSCR) but not found in 2011, using the 1971/81 attrition rates for immigrants:

h_4 =

$$\frac{h_1}{I_1 - k_1}\left(I_4 - k_4\right)$$

Where I_4 = number of LS members entering the study as immigrants in the 2000s (determined by NHS data)

k_4 = number of LS members entering the study as immigrants in the 2000s but having died or embarked by 2011 (determined by NHS data).

Attrition by 2011 of LS members born in the 2000s but not found in 2011, using the 1971/81 attrition rates for those born in the decade, is represented as:

n_4 =

$$\frac{n_1}{B_1 - l_1}(B_4 - l_4)$$

Where B_4 = number of LS members born in the 2000s
l_4 = number of LS members born in the 2000s but having died or embarked by 2011, and n_4 = number of LS members born in the 2000s but not found in 2011.

For 2011, there is also attrition of traced LS members found at the 2001 Census for the first time.

Let M_3 = number of LS members entering and traced in 2001

Attrition by 2011 is represented as

$$\frac{r_1}{N_0 - s_1}(M_3 - d_4)$$

Where d_4 = number of observed 2000s deaths and embarkations for LS members entering by, and traced in 2001.

Thus, 2001/2011 attrition is:

$$\frac{r_1}{N_0 - s_1}\left(\begin{matrix} N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 \\ - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4 \end{matrix}\right)$$

$$+ \quad \frac{h_1}{I_1 - k_1}(I_4 - k_4) \quad + \quad \frac{n_1}{B_1 - l_1}(B_4 - l_4) \quad + \quad \frac{r_1}{N_0 - s_1}(M_3 - d_4)$$

2011/2021 attrition

Attrition for 2011/ 2021 also has four components:

$Attr_5 = r_5 + h_5 + n_5 + m_5$

Here, r_5 represents attrition by 2021 of LS members found in 2011, using 71/81 census-to-census attrition rates:

$r_5 =$

$$\frac{r_1}{N_0 - s_1}\begin{pmatrix} N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 \\ - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4 - r_4 + I_4 - h_4 - k_4 + B_4 \\ - n_4 - l_4 + M_3 - d_4 - m_4 - s_5 \end{pmatrix}$$

s_5 = number of observed 2010s deaths and embarkations for LS members found in 2011.

In addition, attrition by 2021 of LS members entering as immigrants over the 2010s (determined by NHS data) but not found in 2021, using the 1971/81 attrition rates for immigrants:

$h_5 =$

$$\frac{h_1}{I_1 - k_1}(I_5 - k_5)$$

Where I_5 = number of LS members entering the study as immigrants in the 2000s (determined by NHS data)
k_5 = number of LS members entering the study as immigrants in the 2010s but having died or embarked by 2021 (determined by NHS data from Personal Demographic Service).

Attrition by 2021 of LS members born in the 2010s but not found in 2021, using the 1971/81 attrition rates for those born in the decade, is represented as:

$n_5 =$

$$\frac{n_1}{B_1 - l_1}(B_5 - l_5)$$

Where B_5 = number of LS members born in the 2000s
l_5 = number of LS members born in the 2000s but having died or embarked by 2011, and n_5 = number of LS members born in the 2000s but not found in 2011.

For 2021, there is also attrition of traced LS members found at the 2011 Census for the first time.

Let M_4 = number of LS members entering and traced in 2011

Attrition by 2021 is represented as

$$\frac{r_1}{N_0 - s_1}(M_4 - d_5)$$

Where d_5 = number of observed 2010s deaths and embarkations for LS members entering by, and traced in 2011.

Thus, 2011/2021 attrition is:

$$\frac{r_1}{N_0 - s_1}\left(\begin{array}{l} N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 \\ - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4 - r_4 + I_4 - h_4 - k_4 + B_4 \\ - n_4 - l_4 + M_3 - d_4 - m_4 - s_5 \end{array}\right)$$

$$+ \quad \frac{h_1}{I_1 - k_1}(I_5 - k_5) \quad + \quad \frac{n_1}{B_1 - l_1}(B_5 - l_5) \quad + \quad \frac{r_1}{N_0 - s_1}(M_4 - d_5)$$

Link	Attrition component
Based on those present in 1971, adjusted for inter-censal entries and exits	
71/81 t=1	r_1
81/91 t=2	$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2\right)$
91/01 t=3	$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 - m_2 - s_3\right)$
01/11 t=4	$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4\right)$
11/21 t=5	$\frac{r_1}{N_0 - s_1}\left(N_0 - s_1 - r_1 + I_1 - k_1 - h_1 + B_1 - l_1 - n_1 - s_2 - r_2 + I_2 - h_2 - k_2 + B_2 - n_2 - l_2 + M_1 - d_2 - m_2 - s_3 - r_3 + I_3 - h_3 - k_3 + B_3 - n_3 - l_3 + M_2 - d_3 - m_3 - s_4 - r_4 + I_4 - h_4 - k_4 + B_4 - n_4 - I_4 + M_3 - d_4 - m_4 - s_5\right)$
Time t	$\frac{r_1}{N_0 - s_1}\left[\left(N_0 - \sum_{i=1}^{t-1} r_i - \sum_{i=1}^{t} s_i\right) + \left(\sum_{i=1}^{t-1} I_i - \sum_{i=1}^{t-1} h_i - \sum_{i=1}^{t-1} k_i\right) + \left(\sum_{i=1}^{t-1} B_i - \sum_{i=1}^{t-1} n_i - \sum_{i=1}^{t-1} l_i\right) + \left(\sum_{i=1}^{t-2} M_i - \sum_{i=1}^{t-1} d_i - \sum_{i=1}^{t-1} m_i\right)\right]$
Previous decade's immigrants	
71/81 t=1	h_1
81/91 t=2	$\frac{h_1}{I_1 - k_1}\left(I_2 - k_2\right)$
91/01 t=3	$\frac{h_1}{I_1 - k_1}\left(I_3 - k_3\right)$
01/11 t=4	$\frac{h_1}{I_1 - k_1}\left(I_4 - k_4\right)$
11/21 t=5	$\frac{h_1}{I_1 - k_1}\left(I_5 - k_5\right)$
Time t	$\frac{h_1}{I_1 - k_1}\left(I_t - k_t\right)$
Previous decade's births	
71/81 t=1	n_1
81/91 t=2	$\frac{n_1}{B_1 - l_1}\left(B_2 - l_2\right)$
91/01 t=3	$\frac{n_1}{B_1 - l_1}\left(B_3 - l_3\right)$

n_1

01/11 t=4	$\frac{}{B_1 - l_1}(B_4 - l_4)$
11/21 t=5	$\frac{n_1}{B_1 - l_1}(B_5 - l_5)$
Time t	$\frac{n_1}{B_1 - l_1}(B_t - l_t)$
New entrants at the previous census	
71/81 t=1	N/a
81/91 t=2	$\frac{r_1}{N_0 - s_1}(M_1 - d_2)$
91/01 t=3	$\frac{r_1}{N_0 - s_1}(M_2 - d_3)$
01/11 t=4	$\frac{r_1}{N_0 - s_1}(M_3 - d_4)$
11/21 t=5	$\frac{r_1}{N_0 - s_1}(M_4 - d_5)$
Time t	$\frac{r_1}{N_0 - s_1}(M_{t-1} - d_t)$

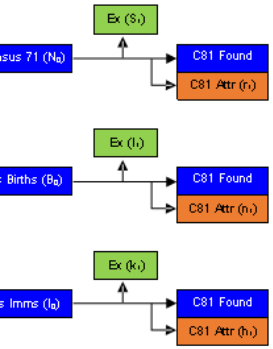
Attrition in time period t can be summarised as:

$$Att_t = \frac{r_1}{N_0 - s_1}\left[\left(N_0 - \sum_{i=1}^{t-1} r_i - \sum_{i=1}^{t} s_i\right) + \left(\sum_{i=1}^{t-1} I_i - \sum_{i=1}^{t-1} h_i - \sum_{i=1}^{t-1} k_i\right) + \left(\sum_{i=1}^{t-1} B_i - \sum_{i=1}^{t-1} n_i - \sum_{i=1}^{t-1} l_i\right) + \left(\sum_{i=1}^{t-2} M_i - \sum_{i=1}^{t-1} d_i - \sum_{i=1}^{t-1} m_i\right)\right]$$

$$+ \frac{h_1}{I_1 - k_1}(I_t - k_t) + \frac{n_1}{B_1 - l_1}(B_t - l_t) + \frac{r_1}{N_0 - s_1}(M_{t-1} - d_t)$$

Attrition rates were calculated separately for different sex/ age groups, so that LS members were grouped into five-year age groups up to 85years, then those aged 85 and over were grouped together.

bserved Attrition 1971-1981



1 k1) Observed Deaths & Embarkations

n rates:

s AtrN $r_1/(N_0-s_1)$

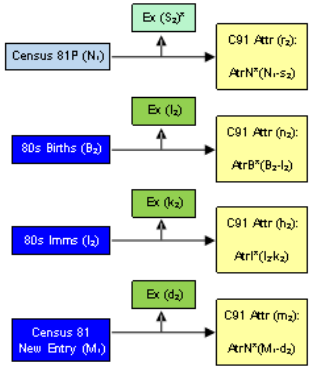
AtrB $n_1/(B_1-I_1)$

ats Atrl $h_1/(I_1-k_1)$

ed present at 1981:

$N_0+B_1+I_1-s_1-i_1-k_1-r_1-n_1-h_1$

Estimated Attrition 1981-1991



Exits:

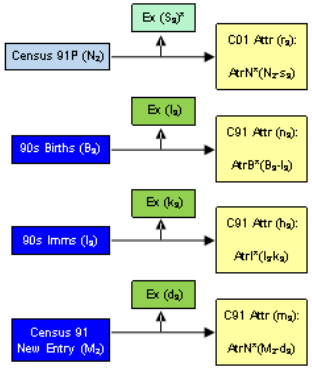
Ex (I_2 k_2 d_2) Observed Deaths & Embarkations

*s_2 N_1 * Exit rate from 1981 enumerations

Projected present at 1991:

N_2 $N_1+B_2+I_2+M_1-s_2-i_2-k_2-d_2-r_2-n_2-h_2-m_2$

Estimated Attrition 1991-2001



Exits:

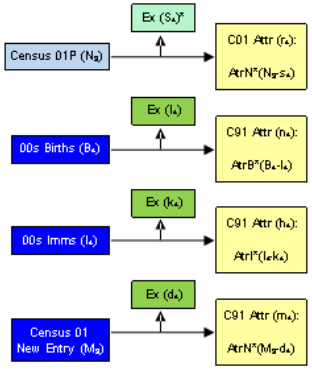
Ex (I_3 k_3 d_3) Observed Deaths & Embarkations

*s_3 N_2 * Exit rate from 1991 enumerations

Projected present at 2001:

N_3 $N_2+B_3+I_3+M_2-s_3-i_3-k_3-d_3-r_3-n_3-h_3-m_3$

Estimated Attrition 2001-2011



Exits:

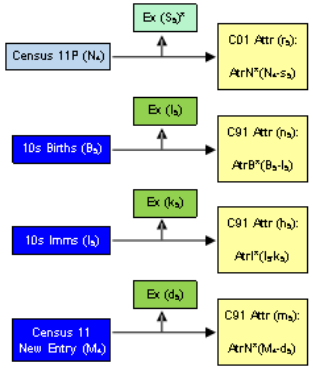
Ex (I_4 k_4 d_4) Observed Deaths & Embarkations

*s_4 N_3 * Exit rate from 2001 enumerations

Projected present at 2011:

N_4 $N_3+B_4+I_4+M_3-s_4-i_4-k_4-d_4-r_4-n_4-h_4-m_4$

Estimated Attrition 2011-2021



Exits:

Ex (I_5 k_5 d_5) Observed Deaths & Embarkations

*s_5 N_4 * Exit rate from 2011 enumerations

Projected present at 2021:

N_5 $N_4+B_5+I_5+M_4-s_5-i_5-k_5-d_5-r_5-n_5-h_5-m_5$