

SARs Use Example 1: Ethnicity

Looking at age, household type and employment with the Household SARs for 1991 and 2001

Jo Wathan, July 2009



This example demonstrates the potential of the Household SAR to describe the household in which a woman lives, to provide context about other household members and to link it with information about her own behaviour.

Some results from 1991 and 2001 are given and the process by which 2001 results were produced are described. By using both years, the range of variables which could be used were limited somewhat. Additionally, as students were treated differently these were excluded once their characteristics had been explored so that the impact of their exclusion could be understood.

The examples shown are:

- Population pyramids by ethnic group showing population age profile changes and the proportions working, not working and full time student
- An example of creating a new variable, which allows us to explore the impact of illness in the household on women's employment across ethnic group
- A complex example of creating a household type classification based on the number of generations in the household to consider the number of extended households within different ethnic groups.

Sample size

The SARs files are routinely large enough to permit individual ethnic groups to be analysed separately. In this example however, we used the smallest SAR to look at subgroups of key ethnic groups. However, this meant that the number of Bangladeshi women in a particular group was sometimes too small to analyse to an adequate degree of precision. Pakistani and Bangladeshi groups were therefore combined once exploratory analysis demonstrated that the characteristics of the groups were similar.

Sample size 2001 SL-Household SAR Female residents aged 16+

White	195,646
Black Caribbean	2,442
Indian	3,910
Pakistani/ Bangladeshi	3,137

Were imputed groups used?

As we know that imputation is unlikely to be accurate at the individual level for ethnicity (e.g. Simpson and Akinwale 2007: 200) exploratory work was undertaken to repeat a series of analyses with and without individuals whose ethnicity was imputed. These cases were identified by the zethew variable; where zethew=1 the

ethnicity variable was imputed for that case. As there was little difference between results including and excluding these individuals the work was undertaken *with* individuals with imputed ethnicities included on this occasion.

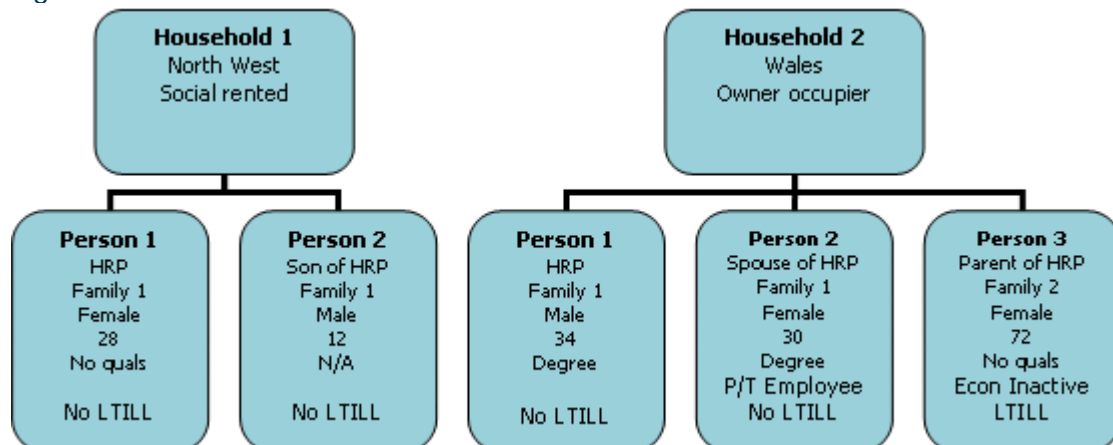
Comparability of Ethnicity in 1991 and 2001

The ethnicity question changed between 1991 and 2001, with the introduction of 'mixed categories'. Further information about the change and its implications is reported elsewhere (see SARs Ethnicity Guide). The decision about which groups to include were driven by the research question, as we sought to extend analyses of groups which are known to differ in their employment patterns. Advice on grouping ethnic categories to maximise comparability was considered (Simpson and Akinwale op. cit., ONS undated). The white group has been shown to be very stable over time. Indian, Pakistani and Bangladeshi groups have been shown to be stable also. The Black Caribbean group is likely to be less consistent between the two time points.

Why was the Household SAR used?

Only by using the household SAR are we able to link household members together. This is because the Household SAR is the only SAR file where entire households are sampled. Each household has a unique ID number which is attached to all individual cases to enable household members to be linked to each other. Because each individual's relationship to the household head is recorded it is possible to create bespoke household type classifications.

Fig 1: the structure of the household SAR



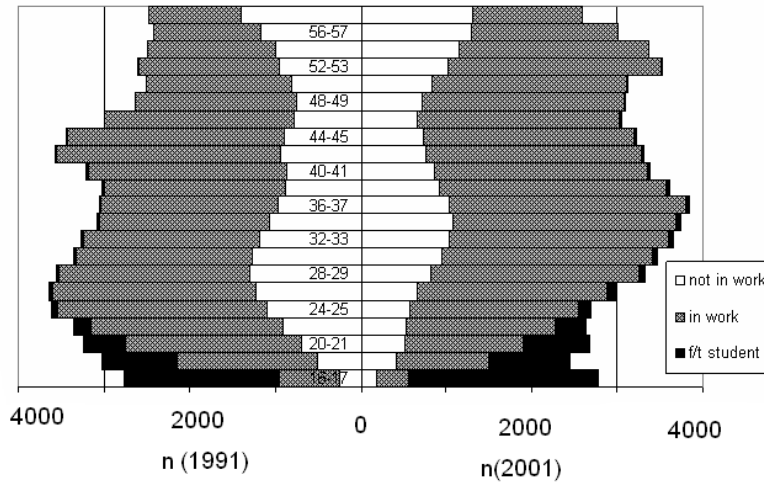
Looking at the relationship between age and employment

White women's employment by age has taken the form of an increasingly shallow M-curve, with the dip associated with absence from the work force while children are young (Dale et. al 2006). An understanding of the age distribution of groups is therefore a helpful starting point for an analysis of the impact of family and household characteristics on women's employment. The charts below are generated from two-way tables of a simple work status variable which distinguishes full time

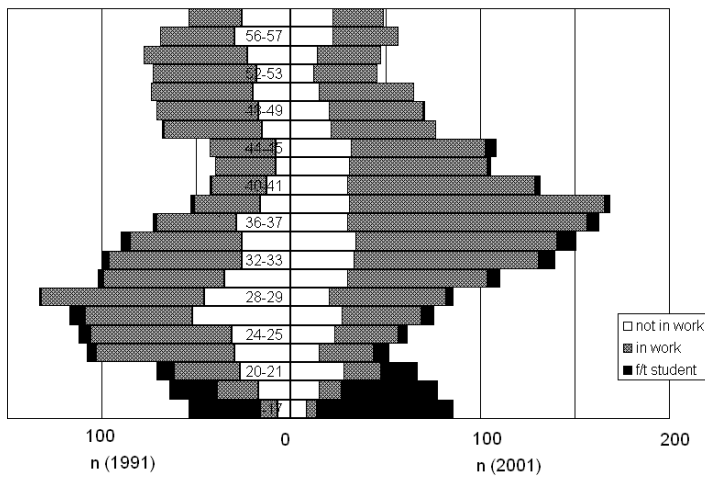
students (prior to being dropped from the data), in paid work and other not in work. We can see clearly what changes in the age distribution of each ethnic group have occurred between 1991 as well as overall levels of paid work by age.

Population Pyramids by ethnic group, Women of working age 1991 and 2001

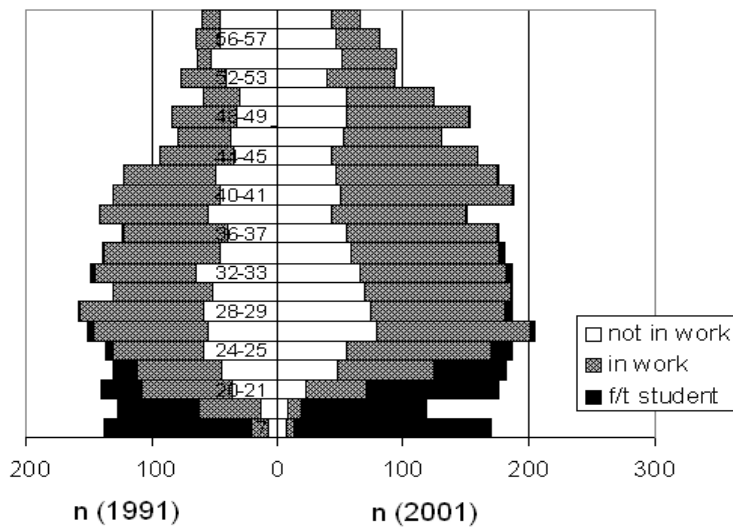
Fig 2a White women (including students)



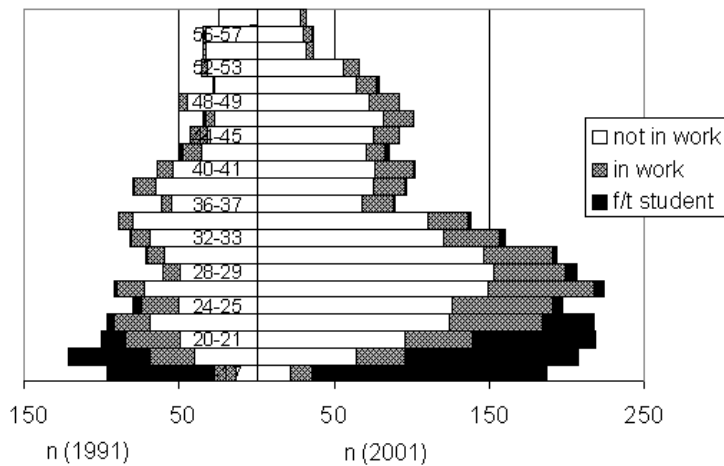
2b Black Caribbean Women



2c Indian women



2d Pakistani and Bangladeshi women



Glancing at these pyramids we can see the impact of ageing on the distribution of the populations. This is most striking in the case of Black Caribbean women, where a bulge in the age distribution peaking at 27-28 in 1991 is echoed in a peak at around age 37-38 in 2001. The relatively even sides of the pyramid for white women are indicative of overall ageing in the population. While the pointed bottom-heavy age distribution amongst Pakistani and Bangladeshi women demonstrates a much younger age profile in both years, but which has grown between 1991 and 2001.

Grey shading dominates most of the pyramids in all but the youngest age groups (where full time students are dominant) for most ethnic groups. The notable exception is the graph for Pakistani and Bangladeshi women where 'not in work' (denoted by the white area nearer the 0 point) is the most common of the three statuses.

Looking at the relationship between illness and employment

Data on the provision of unpaid care was not collected until 2001 so it is not possible to directly explore changes in the impact of caring on women's work. Instead, we

explore whether the presence of a household member with long term limiting illness (LTLI) has an effect on paid work. We might expect:

- Women to work less if they have a long term limiting illness
- Women to be less available for work if they are caring for someone else in the household who has a long term limiting illness

The table below shows the percentage of women of working age who were in paid work for one of four conditions; no household member has LTLI, the respondent has an LTLI, the respondent has no LTLI but other(s) do, the respondent and others do.

Table 1: Percent of women working and count by own and others' long term limiting illness. Female residents aged 16-59(excluding students).

Ethnic group	1991 Household SAR				2001 Special Licence Household SAR			
	None ill	Self ill	Others ill	Both	None ill	Self ill	Others ill	Both
White	71% 103478	26% 5876	62% 11518	31% 3299	79% 95586	36% 11464	69% 16571	30% 5369
Black Caribbean	72% 1258	24% 116	77% 116	34% 47	79% 1333	32% 193	72% 192	34% 70
Indian	63% 1583	21% 106	58% 392	20% 99	69% 1824	30% 256	68% 638	29% 192
Pakistani/Bangladeshi	23% 824	8% 64	20% 292	6% 69	28% 1317	11% 217	29% 638	6% 230

How 'Others ill' was calculated

This was achieved by:

- Ensuring that the long term limiting illness is coded 0 if not ill and 1 if ill
- Summing the long term limiting illness variable across the household (using egen in Stata, or aggregate outfile=* mode=addvariables in SPSS)
- Taking 1 away from the total if the respondent's value of ill equals 1.
- Others are ill if the final total is greater than 0.

The Stata commands used are given in Appendix 1.

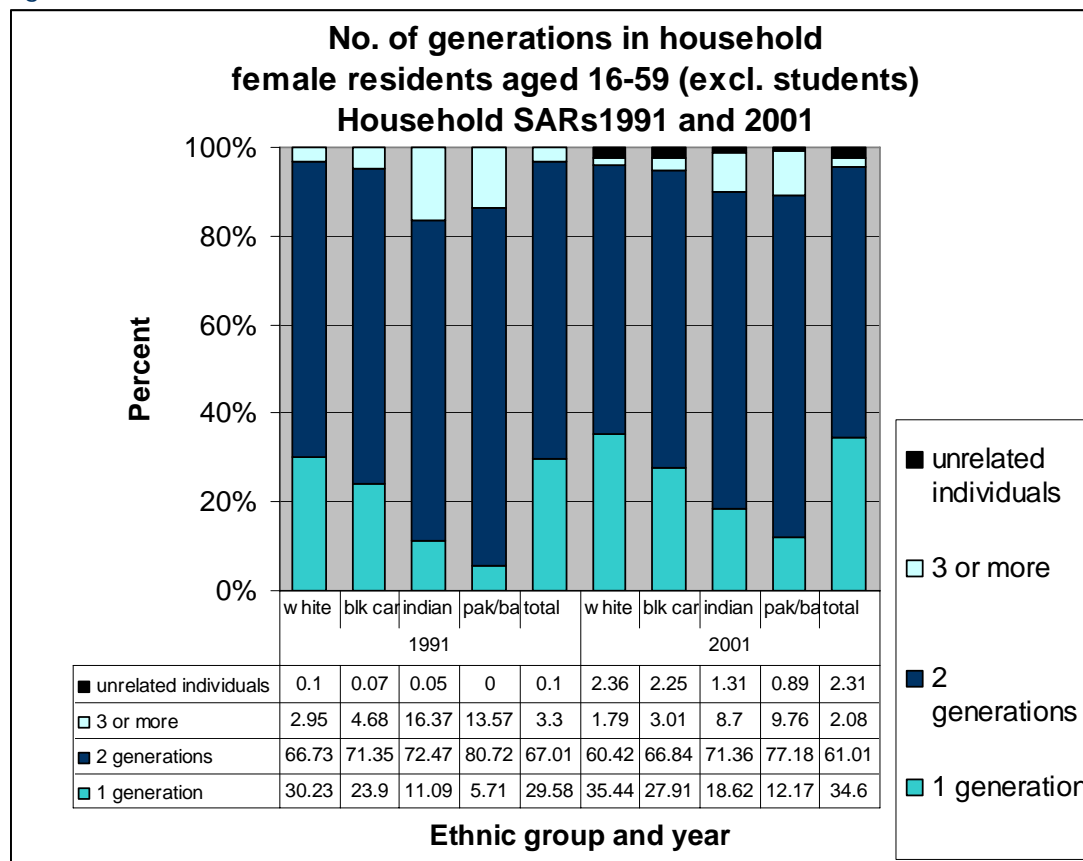
A more complex example: producing a household classification based on number of related generations

Because it is possible for a woman to live in a three (or more) generation family a new variable was created as a step towards identifying which generation a woman was in. This used the relationship to the household reference person, a technique which is not perfect – but allows the identification of three generation households where the members are related to each other.

The approach allowed us to see that:

- Three generation families were least common among white women in both years.
- Three generational families were more common in 1991 than in 2001.
- It was more common for Indian women of working age to live in a three-generational family in 1991 than in a one generation family.
- Households containing unrelated individuals became more common in 2001.

Fig 3



How it was done

The approach was quite involved and had several steps:

1. The relationship to household reference person (as given in the variable reltohr) gives the type of relationship. The relationship can indicate that the person is in the same generation (e.g. spouse, sibling), a higher generation (e.g. parent), one generation below (e.g. child) etc. New variables were calculated to indicate which relative generation the respondent is in.
 - a. Gen1over equalled 1 where the relationship was parent or step parent (reltohr = 7 or 8) and 0 otherwise.
 - b. Gen1less equalled 1 where the relationship was child or step child (reltohr = 4 or 5) and 0 otherwise.
 - c. Gen2less equalled 1 where the relationship was grandchild (reltohr = 9) and 0 otherwise.
 - d. Gen 2over equalled 1 where the relationship was grandparent or step grandparent (reltohr = 10) and 0 otherwise.
2. Once these variables were calculated it was possible to sum the values of each variable across each household. This was achieved using the egen command in Stata (an equivalent approach in SPSS would be to use the aggregate outfile=* mode=addvariables command).
3. These new variables indicate the number of respondents in each of the generations relative to the household reference person. It was possible to determine the number of generations in the household using these:
 - a. The household had 1 generation if all three variables had 0 respondents
 - b. The household had 2 generations if only one of the variables had 1 or more respondents = 1
 - c. The household had 3 generations if 2 or more of the variables had 1 or more respondents = 1

The Stata commands to achieve this are given in the appendix 2.

References and Resources

Dale, A., Lindley, J., and Dex, S. (2006) 'A Life-course Perspective on Ethnic Differences in Women's Economic Activity in Britain' European Sociological Review 22(3) 323-337

Office for National Statistics (undated) A guide to comparing 1991 and 2001 ethnic group data online at <http://www.statistics.gov.uk/articles/nojournal/GuideV9.pdf> <last accessed 07/07/09>

Simpson, L. and Akinwale, B. (2007) 'Quantifying Stability and Change in Ethnic Group' in Journal of Official Statistics Vol. 23 No. 2 pp185-208

SARs Ethnicity Guide <http://www.ccsr.ac.uk/sars/resources/ethnicityguide.pdf> <last accessed 07/07/09>

Guide to Imputation and Perturbation in the SARs (link to be added)

Appendix 1

Stata commands to generate a variable which indicates whether someone other than the respondent has a long term limiting illness.

```
*****health in 1991 we don't have caring***.
*****instead we look to see if self or others in hhd have limiting.
*****long term illness.
ge ill = 0
replace ill = 1 if (llti==1)
ta llti ill, missing
label define ill 0 "not ltill" 1 "ltill"
label values ill ill

****count the no. of ill in the household***.
sort hhid
egen numill = sum(ill), by(hhid)
ta ill numill, missing

ge othill = numill
replace othill = numill-1 if (ill==1)
```

Appendix 2

Stata commands used to generate a variable which counts how many related generations are in the household

```
//*****EXCLUDE STUDENTS
keep if popbase==1

//creating the no. of generations in the family
//household ref person is in reference generation
ge genhoh = 0
replace genhoh = 1 if ((reltohr>=0 & reltohr<4) | reltohr==6)
ge hgen = .
replace hgen=0 if ((reltohr>=0 & reltohr<4) | reltohr==6)
ta reltohr genhoh, missing

//1 gen over if relation to household ref person is parent.
ge genlover = 0
replace genlover = 1 if (reltohr==7 | reltohr==8)
replace hgen = 1 if (reltohr==7 | reltohr==8)
ta reltohr genlover, missing

//1 gen under if relation to household ref person is child.
ge genlless = 0
replace genlless = 1 if (reltohr==4 | reltohr==5)
replace hgen = -1 if (reltohr==4 | reltohr==5)
ta reltohr genlless, missing

//2 gens over if relation to household ref person is g/parent.
ge gen2over = 0
replace gen2over = 1 if (reltohr==10)
replace hgen = 2 if (reltohr==10)
ta reltohr gen2over, missing

//2 gens less if relation to household ref person is g/child.
ge gen2less = 0
replace gen2less = 1 if (reltohr==9)
replace hgen = -2 if (reltohr==9)
ta reltohr gen2less, missing

ta reltohr hgen, missing

//****count the number of generations by looking at number of
// individuals in each relative generation.

ge numgen= 0
replace numgen = 1 if (numgen2over==0 & numgenlover==0 &
numgenlless==0 & numgen2less==0 & numfam==1)
replace numgen = 2 if (numgen2over==0 & numgenlover>0 &
numgenlless==0 & numgen2less==0)
replace numgen = 2 if (numgen2over>0 & numgenlover==0 &
numgenlless==0 & numgen2less==0)
replace numgen = 2 if (numgen2over==0 & numgenlover==0 &
numgenlless>0 & numgen2less==0)
replace numgen = 2 if (numgen2over==0 & numgenlover==0 &
numgenlless==0 & numgen2less>0)
replace numgen = 3 if (numgen2over>0 & numgenlover>0)
replace numgen = 3 if (numgen2over>0 & numgenlless>0)
replace numgen = 3 if (numgenlover>0 & numgenlless>0)
replace numgen = 3 if (numgenlless>0 & numgen2less>0)
replace numgen = 3 if (numgenlover>0 & numgen2less>0)
```