## A Rule-Based Definition for Census Family and Household Reference Persons: Choices and Impacts

Jo Wathan,<sup>\*</sup> Clare Holdsworth,  $^{\dagger}$  and Rachel Leeser<sup>‡</sup>

CCSR Occasional Paper No. 21

ISBN: 1 899005 23 4

\* Centre for Census and Survey Research, University of Manchester

<sup>†</sup> Department of Geography, University of Liverpool

<sup>‡</sup> Greater London Authority

Acknowledgements:

This paper was prepared as part of the consultation exercise undertaken for the Alternative Household Classifications for the 2001 Census project. This project was funded by the ESRC/JISC Census Development Programme (HO507255158). We are grateful to the ESRC for their support and for participants in the consultation exercise for their comments on proposals for a rule-based definition of family and household reference persons.

Material from the Quarterly Labour Force Survey Spring 1996 is Crown Copyright; it has been used been made available by the Office for National Statistics through the Data Archive and has been used by permission. Neither the ONS nor the Data Archive bear any responsibility for the analysis or interpretation of the data reported here.

Thanks are also due to David Rose and David Pevalin at The University of Essex for making available a conversion matrix, which has been used to derive the new NS-SEC classification.

# A Rule-Based Definition for Census Family and Household Reference Persons: Choices And Impacts

### Introduction

The identification of family and household reference persons (FRP and HRP) is an important first step in the preparation of output data for household censuses and surveys. Family and household reference persons are used in two ways:

- 1. to identify an individual person within a household who is used as an anchor or reference point in the production of household derived variables.
- 2. to characterise a household with respect to individual characteristics.

The latter use is perhaps problematic as a single individual can rarely adequately summarise the individual characteristics of a whole family or household. However, it provides a useful, straightforward and common sense way of characterising domestic groupings. In previous censuses the head of household (who for the vast majority of cases was also the head of family unit) was used as the de facto reference person. In 2001 the terminology of the 'head of household' has been dropped and the collection of household matrix data renders the post-hoc identification of a household reference person less problematic. This paper considers possible alternative definitions for a rule-based family and household reference person. Using the household matrix for the 1996 Spring Labour Force Survey for different combinations of rules, including the definition proposed for the 2001 Census, we have compared different options for defining the family reference person. The results demonstrate how the choice of rules impact on the gender profile of family and household references persons. However, as all the proposed definitions rely on some indicator of economic activity, there is very little difference in the economic activity status of representatives identified by the different sets of rules. The paper constitutes background information for those involved with defining family reference people and for secondary analysts who are interested in the impact of a definition change on the meaning of census data and its continuity over time.

## Consultation

A rule-based definition of FRP/HRP was discussed at two consultation workshops where users from academic and local authority backgrounds expressed approval of the idea. In particular, participants at the workshops recognised that a rule-based method of identifying a reference person would introduce a degree of consistency and clarity about the individual identified which would not otherwise be possible.

Participants in the consultation supported using a combination of employment characteristics and age to identify an individual whose personal characteristics were most likely to influence the life chances of other members. It is intended that this person should closely correspond to be the family or household member who has the greatest economic impact on the household or family (and is thus likely to be the chief income earner, see Market Research Society, 1991).<sup>1</sup> In most cases this person will also be the individual formerly identified as the head of household.

<sup>&</sup>lt;sup>1</sup> The chief income earner is the person with the largest income, whether from employment, pensions, state benefits, investments or any other source, for a household group of related individuals.

## Methodology

We therefore developed and tested a number of different rules for defining the family and household reference person based on the following general principles:

- 1. The method should be easy to describe in small print and readily understandable. No benefit is to be had if the rules are not well known by Census users.
- 2. The method should be easy to program.
- 3. The family reference person should be identified first. In the case of a household consisting of two or more family units the household reference person should either be chosen by applying the same rules used to identify the family reference person, or by selecting the oldest family reference person. The former is the agreed method for selecting the household reference person for the 2001 Census. However, in households containing more than one family group, family heads of 'proper families' (i.e. families with two or more members) will always take precedence over 'ungrouped individuals' (families of one person only). For example, in the case of a couple living with an elderly parent, only the family head of the couple family would be considered as the household reference person.
- 4. The family reference person should be an individual in the parental generation of the family unit. This limits any selection of FRP to one of a maximum of two people.
- 5. Any rule must ultimately rely on person number on the census form as this is the only variable for which there will be no tie between two individuals. In many cases the tie will involve the first person on the form and another. As the first person on the form should be 'the householder', this method will give the required result.
- 6. The upper generation in the family unit is constituted by person1 or person2. Person 1 is the first of the couple to be entered onto the form, person 2 is the second. It is not assumed that person 1 is male, or that there will be one male and one female.

We identified two specific approaches to defining a rule-based definition of FRP:

- 1. The first is combinatorial, based on the following criteria:
  - Working/not working status
  - Hours worked
  - Occupation: based on a 3-fold category of NS-SEC (for current and previous workers)
  - Year last worked (if not currently in work)
- 2. The second is based on full-time and part-time working, and treats the unemployed and pensioners separately.

NS-SEC was used at the 3-category level as this is the only level at which the classification is even arguably ordinal (Rose and O'Reilly 1998). This classification was recreated using a matrix provided by David Pevalin and David Rose which enabled NS-SEC to be replicated from a combination of SOC-90, Employment status, managerial/supervisory status and size of institution.

Hours are used rather than a full/part/no work tripartite classification. Hours provide greater ability to distinguish between individuals (although the distinction might be spurious, longer hours tend to be worked by those with greater earnings).

## Rule definitions: Approach 1 combinations of criteria

As the family reference person is defined first, we have concentrated on the definition of FRP rather than HRP. Two different sets of rules were created and combined;

- four rules for families with someone in the upper generation in work
- six rules for families with no one in the upper generation in work. In the 1996 LFS just under 15% of households recorded no one in the upper generation in work.

In each rule combination the individual described is taken to be the family reference person at each numbered stage. In the event of a tie, or the classification failing to be defined, the next stage is considered. So, for example, under algorithm frpw1 the sole working adult in the upper generation would be taken as the family reference person. If this criterion did not identify a unique individual within the family unit, criterion 2 would be applied, and so on.

For families in which at least one person in the upper generation is in work the following sets of alternative rules were considered and the results compared:

## frpw1

- 1. sole adult in upper generation
- 2. sole worker in upper generation
- 3. the worker with the longer hours in upper generation
- 4. the worker with the 'highest' NS-SEC in upper generation
- 5. the first individual on the form in the upper generation

## frpw2

- 1. sole adult in upper generation
- 2. sole worker in upper generation
- 3. the worker with the longer hours in the upper generation
- 4. the worker with the 'highest' NS-SEC in upper generation
- 5. the oldest worker in the upper generation
- 6. the first individual on the form in the upper generation

## frpw3

- 1. sole adult in upper generation
- 2. sole worker in upper generation
- 3. the worker with the 'highest' NS-SEC in the upper generation
- 4. the worker with the longer hours in the upper generation
- 5. the first individual on the form in the upper generation

## frpw4

- 1. sole adult in upper generation
- 2. sole worker in upper generation
- 3. the worker with the 'highest' NS-SEC in the upper generation
- 4. the worker with the longer hours in the upper generation
- 5. the oldest worker in the upper generation
- 6. the first individual on the form in the upper generation

In the event that there is no worker in the family the following rules were applied and tested:

## frpn5

1. the first individual on the form in the upper generation

## frpn6

- 1. the oldest person in the upper generation
- 2. the first individual on the form in the upper generation

## frpn7

- 1. the person who was most recently in work in the upper generation
- 2. the first individual on the form in the upper generation

## frpn8

- 1. the person who was most recently in work in the upper generation
- 2. the oldest person in the upper generation
- 3. the first individual on the form in the upper generation

## frpn9

- 1. the person with the higher NS-SEC for their most recent job in the upper generation
- 2. the first individual on the form in the upper generation

## frpn10

- 1. the person with the higher NS-SEC for their most recent job in the upper generation
- 2. the oldest person in the upper generation
- 3. the first individual on the form in the upper generation

Flowcharts illustrating the decision making process for each of these rules (frpw1 to frpn10) are given in the appendix.

We paired together different combinations of rules 1-4 and 5-10 to cover all families, whether or not they contain anyone in work. Not all of the combinations of rules are viable; we identified 18 possible combinations (table 1) and used the Spring1996 Labour Force Survey to compare the characteristics of the family reference person identified for each combination.

A slight difference between LFS and Census data is the availability of information on previous job. This is only available for LFS respondents who have worked in the 8 years preceding the survey date. This is therefore more limited than in the census. Another difference between the Labour Force Survey and the Census is that the form is not self-completion questionnaire and therefore there is zero possibility within the LFS that a baby, for example, would be recorded as person 1, indeed person 1 is normally head of household. This contrasts with the Census and affects our ability to compare the selection rules with those applied in the 1991 Census.

Combination Number	Rule if at least one	Rule if no member of the			
	member of the upper	upper generation is in			
	generation is in work	work			
1	Frpw1	Frpn5			
2	Frpw1	Frpn7			
3	Frpw1	Frpn9			
4	Frpw2	Frpn5			
5	Frpw2	Frpn6			
6	Frpw2	Frpn7			
7	Frpw2	Frpn8			
8	Frpw2	Frpn9			
9	Frpw3	Frpn10			
10	Frpw3	Frpn5			
11	Frpw3	Frpn7			
12	Frpw4	Frpn9			
13	Frpw4	Frpn5			
14	Frpw4	Frpn6			
15	Frpw4	Frpn7			
16	Frpw4	Frpn8			
17	Frpw4	Frpn9			
18	Frpw4	Frpn10			

Table 1: Combinations of rules used to define Family Reference Person

The classification is calculated for all family units as defined in the LFS. This includes persons who will be returned as 'ungrouped individuals' (defined here as one-person family units) in the census. This group includes one-person households, but also individuals in shared households, such as a student household or siblings living together, where each individual household member is classified as an 'ungrouped individual'. This permits the classification to be used to produce tables characterising all households, not just those containing a couple or lone-parent family. As a result, however, there may be a conflict between the definition of family as used in this paper and that used in the Census. Ungrouped individuals (i.e. families of size one) compose 15% of the population and 35% of all families<sup>2</sup>. Their inclusion has a consistent effect on the results given here as in each case the family reference person will be defined as person 1. Their characteristics are given, for reference, in row 19 of table 2.

After applying each of the rule combinations to the 1996 LFS we have compared the following characteristics of FRPs (table 2):

- 1. Percentage of Family Reference Persons who are listed as having the lowest person number (i.e. FRP is the first person in the upper generation of the family listed in the data).
- 2. Percentage of Family Reference Persons who are households heads in the LFS.
- 3. Percentage of Family Reference Persons who are male.
- 4. Percentage of Family Reference Persons who are economically active.
- 5. Mean net pay of the Family Reference Person (given for those in work only).

The closest possible approximation for the 1991 Census definition of Head of Household is to take the first person in the form. The values of the characteristics of

<sup>&</sup>lt;sup>2</sup> Calculated from the Quarterly Labour Force Survey Spring 1996.

the first person are given in the final row of table 2. As already noted, the Labour Force Survey has a different methodology for allocating person number which is likely to produce a greater number of adults as the first person on the form.

	Col 1	Col 2	Col 3	Col 4	Col 5
Combination	% <b>FRP</b> =	%	%	%	Mean net
	lowest person	LFS Head	male	econ active	pay of FRP
	number	01 Household			(main job, workers
		nousenoiu			only)
1. Frpw1 & frpn5	92	85	64	62	234
2. Frpw1 & frpn7	89	82	61	62	
3. Frpw1 & frpn9	91	83	63	62	
4. Frpw2 & frpn5	88	80	60	62	220
5. Frpw2 & frpn6	85	78	57	62	
6. Frpw2 & frpn7	84	77	57	62	
7. Frpw2 & frpn8	84	77	56	62	
8. Frpw2 & frpn9	86	79	58	62	
9. Frpw2 & frpn10	86	78	58	61	
10. Frpw3 & frpn5	88	81	60	61	228
11. Frpw3 & frpn7	85	78	57	61	
12. Frpw3 & frpn9	87	80	59	61	
13. Frpw4 & frpn5	84	77	56	61	216
14. Frpw4 & frpn6	81	74	54	61	
15. Frpw4 & frpn7	81	74	53	61	
16. Frpw4 & frpn8	81	74	53	61	
17. Frpw4 & frpn9	83	75	55	61	
18. Frpw4 & frpn10	82	75	54	59	
19. Characteristics of	-	82	43	43	210
ungrouped					
individuals					
20. Characteristics of	-	99	74	61	249
the person 1 on					
the LFS schedule					

Table 2: Characteristics of Family Reference Person for each Combination of rules

Source: Labour Force Survey Spring Quarter 1996 Crown copyright

An appropriate algorithm for identifying FRP should have a high percentage in the first two indicators (columns 1 and 2), indicating comparability with family reference person and head of household respectively. This is desirable from the point of view of continuity over time. Moreover, a high percentage of economically active FRPs and high mean net pay would suggest a high degree of correspondence between FRP and Chief Income Earner, which would satisfy a second aim – that of identifying an individual whose own characteristics were most likely to impact on the life chances of his or her family. However, these two aims may be contradictory: increasing the chances of finding higher earners and individuals who are more likely to be

economically active may lead to a lower correspondence between the previous methods of identifying head of household.

The first column of table 2 demonstrates that, for all definitions, the order in which individuals appear on the schedule will be strongly related to their chances of being identified as the family reference person. Between 81% and 92% of family reference people are either the sole adult, sole adult in the parental generation or first of two adults in the generation listed on the schedule.

For the 18 algorithms considered here, between 74% and 85% of family reference persons were also defined as Household of Household according the 1996 LFS definition (Column 2, Table 2)<sup>3</sup>. In some cases, this difference can be attributed to the mismatch of family and household, as 6% of households contain two or more family units. However, this effect will be consistent across all definitions, and variations between different algorithms indicate the extent to which each rule combination overlaps with the LFS 1996 definition of Head of Household. The LFS definition relates to responsibility for housing, but has an intrinsic male bias, which means that any woman married to or cohabiting with a man cannot be defined as head of household. Accordingly, the rule generally only distinguishes between individuals on grounds *other* than sex. across different family units.

We may use columns 1 and 2 of table 2 as indicators of the degree of divergence with previous definitions of head of household. The more complex algorithms diverge to a greater extent than the simpler algorithms. The lowest percentages in columns 1 and 2 are found when either frpw2 or frpw4 are used; both of these rules include a control on age. Combinations including frpw4 are particularly unlikely to produce family reference people coincident with the Labour Force Survey definition of head of household, or to identify the first adult in the family on the schedule. Frpw4 prioritises NS-SEC over hours worked and includes age as a tie-breaker criterion, whereas frpw2 prioritises hours worked over NS-SEC. It is clear that the order of rules as well as the inclusion of rules influences the choice of FRP.

The percentage of FRPs who are male ranges from 53% to 64% depending on the rule used. The lowest person number rule and the LFS definition of head of household both identify individuals who are more likely to be male any of these rules. This characteristic is unsurprising; the LFS head of household definition has an intrinsic male bias. We find the greatest discrepancy in the sex of the FRPs between rule combinations based on frpw1 and frpw4. The difference between these two classifications is that frpw1 prioritises hours over NS-SEC and does not include a control for age, while for frpw4 NS-SEC is taken before hours and a control for age is included before taking the first person on the form. This latter combination may therefore define women working part-time in higher NS-SEC occupations as FRPs.

In all cases, criteria used to identify an FRP are based on variables that are applicable only to economically active individuals. Accordingly, we find that there is no relationship between the economic activity rate of the FRP and the algorithm used and all algorithms are as likely to identify an economically active person as simply taking the first person on the form.

<sup>&</sup>lt;sup>3</sup> The head of household (HOH) is defined in 1991 as either the man or the husband/male partner of the woman in whose name the accommodation was owned or rented. Where two people have equal claim the either the oldest male is selected or, in all female households, the oldest female (McCrossan 1991).

All algorithms are less likely to identify the highest earner as FRP than simply taking the first person in the household. In particular, including age appears to lower the chances of identifying an FRP with high earnings It should be noted that pay as given here refers to the net weekly pay of workers in their main job and that individuals who do not work are not included in the calculation of the mean. This finding is contrary to one of the main aims of using a rule-based definition of FRP, which is that the rule should attempt to identify the Chief Income Earner. However, it should be noted that the procedure to define the first person in the LFS schedule is different from that used in the Census. In the LFS the first person on the schedule will be (in practically all cases) the Head of Household as identified by the interviewer.

Finally, the overall impact of changing the selection rule to select between two individuals who are not working is less than that for the rules for individuals in work. This is consistent with the limited applicability of these criteria to workless families only and with the restriction of the selection criteria choice of either including an age rule or not. Clearly this part of the rule does not impact on pay or economic activity (subject to rounding error) as the criteria only applies once economic activity has been taken into consideration.

#### **Rule definitions: Approach 2**

The second method treats individuals who are recently unemployed (having left a previous job in the same calendar year<sup>4</sup>) as still having the same role in the family as a person in work. This assumes that the individual is temporarily unemployed and that the income and of this individual of a temporarily unemployed person may still be most likely to impact on the life chances of the rest of the family. Additionally it includes a control between full and part-time work instead of taking usual hours of work. Full time work is defined as working more than 30 hours per week<sup>5</sup>

The order of priority used was:

- 1. The sole current or recent worker in the upper generation.
- 2. The sole full-time worker in the upper generation.
- 3. The current or recent worker with the 'higher' NSSEC in the upper generation.
- 4. The sole unemployed individual in the upper generation (excluding recent unemployed).
- 5. The sole retired upper generation person in a workless family.
- 6. The first upper generation person on the form.

This algorithm resulted in the identification of an FRP with the following characteristics:

- in 99% of cases the FRP is the first person on the form
- in 91% of cases the FRP was the Household head in the LFS
- FRPs are male in 70% of cases
- FRPs are economically active in 59% of cases
- FRPs have a mean net weekly pay of 245 pounds in the main job.

<sup>&</sup>lt;sup>4</sup> The questions asked in the census mean that is will be defined in terms of work during the current calendar year rather than in the last twelve months in the Census.

<sup>&</sup>lt;sup>5</sup> Within the census the self-defined distinction can be used to determine whether an individual is working full or part-time; there is no such distinction in the LFS.

This method therefore shows much greater consistency with previous definitions for FRPs and HRPs (i.e. taking the first person on the Census form), and a higher proportion of men, though a lower proportion of economically active FRPs. This latter characteristic is due to the priority given to retired members of the households. The absence of a control for age may account for the larger proportion of male FRPs, though the inclusion of unemployed in the algorithm, prior to first person on the form may also increase the proportion of men (reflecting gender differences in unemployment rates).

### **2001 Census Definition**

The 2001 Census will use the following definition of FRP, which is based on categories of economic activity, age and person number on the census form:

- 1. Sole adult in the upper generation
- 2. Sole worker in full-time job in upper generation who is not an economically active full-time student.
- 3. Sole worker in full-time job in upper generation who is an economically active full-time student.
- 4. Sole worker in part-time job in upper generation who is not an economically active full-time student.
- 5. Sole worker in part-time job in upper generation who is an economically active full-time student.
- 6. Sole unemployed person in the upper generation who is not an economically active full-time student.
- 7. Sole unemployed person in the upper generation who is an economically active full-time student.
- 8. Sole retired person in the upper generation.
- 9. Sole other economically inactive person in the upper generation.
- 10. Oldest person in the upper generation
- 11. The first person on the census form in the upper generation.

If there is more than one family unit in the household, the same criteria will be used to select the Household Reference Person from the FRPs. However, only FRPs of 'proper families' (i.e. families with two or more members) will be considered for selection of the Household Reference Person. In households consisting of two or more 'ungrouped individuals' only (i.e. no proper families), the above rules will be used to select the HRP.

This algorithm resulted in the identification of an FRP with the following characteristics:

- in 90% of cases the FRP is the first person on the form
- in 83% of cases the FRP was the Household head in the LFS
- FRPs are male in 62% of case
- FRPs are economically active in 63% of cases
- FRPs have a mean net weekly pay of 236 pounds in the main job.

Although this at first appears to be more similar to Approach 2 than to the combinatorial approach, it would appear that the rule proposed for identifying the family reference person in the census produces results similar to the combinatorial methods which prioritised hours of work, then NS-SEC (i.e. those based on frpw1).

### Conclusions

It is not possible using Labour Force Survey data alone to fully assess the extent to which a rule based definition of FRP for the 2001 Census will differ from that of the previous selection criterion of taking the first person on the form. However, we have shown that for survey data, rule based methods may produce a higher proportion of female FRPs and can identify FRPs with a high rate of economic activity. However it is unclear from the analysis of the LFS is this will necessarily mean that FRPs identified using a rule-based approach will have higher earnings. Finally, the methods considered here have the benefit of being clearly defined, which diminishes doubt as to the meaning of the reference person category.

Although the methods vary somewhat, the characteristics of the FRPs identified are reasonably consistent with respect to four indicators. None of these indicators differ by more than ten percentage points across the different combinations of rules considered.

None of the rules used involved complex programming, the greater part of the SPSS syntax used involved the conversion of the SOC90 to NS-SEC. While the length of some of the rules are longer than might be ideal from a point of view of disseminating metadata, the principles behind each rule is straightforward and understandable.

## References

Market Research Society (1991) Occupational Groupings: A Job Dictionary, Market Research Society, (London), p 5.

McCrossan (1991) A Handbook for Interviewers OPCS Social Survey Division, HMSO (London)

Rose, D. and O'Reilly, K. (1998) *The ESRC review of government social classifications* Office for National Statistics (London)



FRPW1: Version 1 Family Reference Person (where at least one partner is in work) Document F3  $\,$ 











FRPW4: Version 4 Family Reference Person (where at least one partner is in work) Document F6 : Upper generation people only



FRPN5: Version 1 Family Reference Person (where no partner is in work) Document F7



FRPN6: version 2 Family Reference Person (where no uppergeneration person is in work) Document F8



FRPN7 version 3 Family Reference Person (where no member of the upper generation is in work) Document F9







FRPN9 version 3 Family Reference Person (where no member of the upper generation is in work) Document F11



FRPN10 version 3 Family Reference Person (where no member of the upper generation is in work) Document F12