

Samples of Anonymized Records (SARs) from the UK Censuses: A Unique Source for Social Science Research

■ **Yaojun Li**

Birmingham University

ABSTRACT

The Samples of Anonymized Records (SARs) from the 1991 UK census have proved an outstanding achievement for social science research. Research based on the SARs has taken advantage of the large sample size, the detailed geography, the wide range of socio-demographic variables and the operational flexibility. A wide range of areas has been covered including sociology, human geography and public health. The new SARs from the 2001 UK census will provide an even richer source of information. The availability of the SARs from the two censuses will provide a unique opportunity for studying socio-economic cultural and demographic changes in the UK over the decade and for conducting international comparisons. This article reviews the data structure of the SARs, summarizes some of the most innovative and important findings using the 1991 SARs, and looks at the research potential of the 2001 SARs. The purpose of the article is to encourage even greater use of the SARs.

KEY WORDS

class / education / employment / ethnicity / religion / SARs

Introduction

Marsh and her colleagues made the case for a Sample of Anonymized Records (SAR) from the 1991 UK census of population (Marsh et al., 1991). The case was accepted by the Office for National Statistics (ONS, then the Office of Population Censuses and Surveys) and a request was

made by the Economic and Social Research Council (ESRC) to purchase the SARs. A 2% Individual SAR and a 1% Household SAR were released for Great Britain. Similar samples were subsequently released for Northern Ireland. The Cathie Marsh Centre for Census and Survey Research (CCSR, formerly Census Microdata Unit), based at Manchester University, was set up in April 1992 for the dissemination and support of the SARs. Since their release, the SARs have proved an enormous success for social science research in the UK and abroad. Users of the 2% Individual SAR have exploited the large sample size and relatively detailed geography to look at differences between sub-populations (especially ethnic groups) and between geographical areas. Users of the 1% Household SAR have exploited the hierarchical nature of the data to study various aspects of household and family structure and interrelationships within households. Research using the SARs has covered a wide range of disciplines such as sociology, human geography, public health, social policy and social statistics. Around 400 papers have been published. Many of these, written by some of the best-known social scientists, contain highly important substantive findings and methodological innovations.

The advent of the SARs from the 2001 UK census will present even greater opportunities. The new SARs will retain similar structures to the 1991 SARs but will have new or more detailed variables, making them an unrivalled source for studying the various aspects of social life in the UK at the turn of the century, for exploring the social changes between the two censuses, and for conducting international comparisons.¹ This article gives a brief introduction to the structure of the SARs, reviews some of the most important SAR-based findings within the sociological remit, and shows how the 2001 SARs can be used both on their own and in combination with the 1991 SARs for comparative research.² The purpose of the article is to encourage even greater use of the SARs.

The Structures of the SARs

SARs from the 1991 Census

The 2% Individual SAR is a sample of individuals enumerated in households and communal establishments. It has 1.2 million records in 12 regions and 278 geographical (SAR) areas. The SAR areas include all large local authority districts with a population of at least 120,000 in the 1989 mid-year estimates. Smaller local authorities under this threshold have been grouped for confidentiality reasons. As shown in Figure 1, each record has a personal id and 66 variables at three levels: household, family and individual. Household level variables contain information shared by all members of the household such as number of cars in household, housing tenure, and amenities like bath or central heating. Family level variables concern family situations such as family type (famtype) or social class position of the family head (sclassfh). Individual level

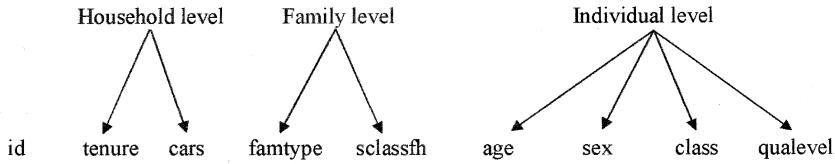


Figure 1 Data layout of the 2% SAR

variables cover a wide range of socio-economic cultural and demographic data such as age, sex, marital status, ethnicity, employment status, levels of educational qualifications (qualevel), social class, and health conditions. These variables have proved most useful for sociological research.

The 1% Household SAR is a hierarchical file which allows linkages between individuals within family and household. The geographical base is the Registrar General's Standard Regions in 1991, with the South East split into Inner London, Outer London and the remainder of the South East. The file contains 215,789 households, 153,491 families and 541,922 individuals (89,254 people are not in families). Although 98.73 percent of the households in the file have only one family, 1,943 households contain 2 to 4 families. There are about 40 original variables in the file, similar to those on the Individual SAR. However, the structure of the file allows a large number of variables to be derived, and around 80 such variables have been added. Reflecting the questions in the census schedule, there is a set of variables for the household and a further set of variables for each individual within the household. Thus each individual in the file has not only his or her personal information such as age, sex, ethnicity, marital status, class, education and employment status, but also household level data such as housing tenure and number of cars in the household. Because of the hierarchical nature of the file, it has proved a very helpful source for teaching and research.³ The hierarchical structure of the file is shown in Figure 2. In this example we have a household of five people, who belong to two families. The different research values of the Individual and the Household SARs can be noted here. While the former has a lot of detailed individual and geographical information which enables research of considerable academic and policy importance, the latter is suitable for studying relationships within households such as 'work-rich'/'work-poor' households, or the ethnic composition of the household. The next section will give examples of the research value of the two files.

SARs from the 2001 Census

The structure of the SARs from the 2001 census is similar to that of the 1991 SARs. There are, however, some notable differences in the 2001 census schedule, and hence in the variables of the 2001 SARs. The differences lie

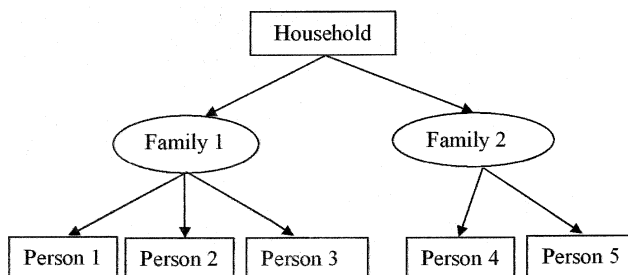


Figure 2 The hierarchical structure of the 1% SAR

mainly in the new class schema adopted, greater differentiation of educational and professional qualifications, religious affiliation (for the whole of UK) and religious background (for Scotland and Northern Ireland), general health and caring questions, the household membership matrix, and the changing definition of unemployment from self-reported unemployment in the 1991 census to the ILO (International Labour Organisation) definition in the 2001 census.⁴ Because of their particular research potential for sociology, we give a brief discussion of three of the most important changes: class, education and religion. The other changes are explained, where appropriate, in the discussion later.⁵

In the 1991 SARs, two class schemas were provided: the Registrar General's social class (SC) and the Social Economic Groups (SEGs). These schemas, although much used in research, are criticized for lacking 'a clear conceptual basis' (Rose and O'Reilly, 1998: 31). As a result of the ESRC Review of Social Classification, a new class schema, called National Statistics Socio-Economic Classification (NS-SEC), has been adopted for the 2001 census. This schema, based on the employment relationship theory (Erikson and Goldthorpe, 1992; Goldthorpe et al., 1987), is close to the Goldthorpe class schema widely regarded as the best of all existing sociological endeavours (Evans and Mills, 2000; Marshall et al., 1988; Rose and O'Reilly, 1997, 1998; Rose and Pevalin, 2001). The theory is drawn from Marxist and Weberian origins, and aimed chiefly at differentiating between the 'service' employment relationship of professionals and managers typified by current rewards, future compensations and embedded career structures on the one hand, and the labour contract relationship of the working class entailing a relatively short-term and specific exchange of money for effort on the other (Erikson and Goldthorpe, 1992: 35–47).

The NS-SEC (Figure 3) represents a variety of labour market positions and employment statuses, and covers all the adult population. The schema has two levels of differentiation. First, it differentiates between employment statuses: employers, self-employed and employees. Second, for the majority of occupations within the employee status, it further differentiates higher and lower grades of professional and managerial positions, intermediate positions, lower

Full Version		Nine class		Eight class		Five class		Three class	
L1	Employers (large)	1A	Higher managers	1.	Higher managerial & professional	1.	Managerial and professional	I	Managerial and professional
L2	Managers (large)		1B						
L3	Professionals	2.		Lower managerial & professional					
L4	Associate professionals								
L5	Managers (small)	2.	Lower managerial & professional	2.	Lower managerial & professional	2.	Intermediate	II	Intermediate
L6	Higher supervisors		3.						
L7	Intermediate employees	4.		Small employers and own account	4.	Small employers and own account	3.	Small employers and own account	
L8	Employers (small)		5.						Supervisor/craft related
L9	Own account	6.		Semi-routine employees	6.	Semi-routine employees	5.	Working class	
L10	Lower supervisors		7.						Routine employees
L11	Craft and related employees	8.		Never worked/long-term unemployed	8.	Never worked/long-term unemployed			
L12	Semi-routine employees								
L13	Routine employees								
L14	Never worked/long-term unemployed								

Figure 3 The National Statistics Socio-Economic Classification (NS-SEC)

Source: Rose and O'Reilly, 1998: 22, Figure 6. National Statistics website: www.statistics.gov.uk. Crown copyright material is reproduced with the permission of the Controller of HMSO

1991 SARs	2001 SARs	Notes
SOCLASS*	SOCLASS	Registrar General's class
SEGROUP*	SEGROUP	Socio-Economic Groups
GCLASS [†]		Goldthorpe class
NS-SEC	NS-SEC*	National Statistics Socio-Economic Classification; 5-category for 1991 SARs
CAMSCORE		Cambridge Occupational Scores
CAMSCORE	CAMSIS	Cambridge Social Interaction and Stratification Scale
ISCO [†]	ISCO [†]	International standard classification of occupations
ISEI [†]	ISEI [†]	International Socio-Economic Index of Occupational Status
SIOPS [†]	SIOPS [†]	Standard International Occupational Prestige Scores

Notes:

- 1 Variables marked by * are supplied by the ONS and the rest are derived by CCSR in consultation with other researchers and added to the files. For details of ISCO, see <http://www.fss.uu.nl/soc/hg/isko88>; for details of CAMIS, see <http://www.cf.ac.uk/socsi/CAMSIS/index.html>
- 2 All variables, regardless of the mark, are available in the 1% SAR at the household, family and individual levels. For instance, soclass is available for the individual, the family and the household reference persons.
- 3 Variables marked by [†] are not available on the 2% SARs.

Figure 4 Class measures in the SARs

supervisory and technical positions, and semi-routine and routine positions. Depending on the research purposes and sample sizes, the categories can be collapsed into nine, eight, five and three classes.

In order to provide the continuity for research, all the 'class' measures in the SARs will be standardized. First, NS-SEC will be derived for the 1991 SARs. Second, all class variables available in the 1991 SARs will be added to the 2001 SARs (see Figure 4). This will ensure that the SARs will have the same types of class variables.

Educational qualifications in the 1991 SARs are available for people aged 18 or over. Three levels of education are recorded for people with post-secondary education: higher degrees, first degrees and professional qualifications below degree level such as teaching and nursing (OPCS, 1992: 44–5). In the 2001 SARs, not only will the age threshold be lowered to 16, but the question refers to the full range of qualifications as well. For England and Wales, various professional qualifications will be available such as teaching, nursing, midwifery or dentist qualifications. The greater differentiation of educational and professional qualifications will be of important value to both academic and policy research, as seen later.

The 1991 census contained no religion questions for Great Britain. Thus, in spite of the availability of ethnicity and educational qualifications, research into the cultural facets of life was limited. This situation will change with the 2001 SARs, as questions of religious affiliation (for the whole of UK) and of religious background (for Scotland and Northern Ireland) are available in the 2001 census.

Information about access, training and support for the SARs is available online.⁶ Accessing is easy. After registration, the 1991 and the 2001 SARs can be directly downloaded in various formats such as SPSS or Stata. CCSR will organize ‘road-shows’ and training sessions across the country. CCSR also offers a variety of day courses, many of which are aimed specifically at the use of the SARs.⁷ SAR-related expert support is available from the website with feedback guaranteed within two working days.

Research Potential of the SARs: Examples and Suggestions

As mentioned earlier, the large sample size, the relatively detailed geography, and the availability of a variety of socio-economic-demographic variables in the 1991 SARs have enabled research not normally feasible using social surveys. In the following, we select a few themes from publications based on the 1991 SARs and make suggestions for further research with the 2001 SARs. As so much has been published, we do not claim to give a full or representative summary of all the findings from existing research. The following provides some examples that are both important and relevant for sociological research. It is organized into six areas: ethnicity, education and class; employment and the labour market; religion; health and caring; household and family composition; and the use of SARs for research in the regional (Scotland, Wales and Northern Ireland) context and for international comparisons.

Ethnicity, Education and Class

Prior to the release of the 1991 SARs, research into ethnic differences in the UK had been hampered by the sample sizes of ethnic groups in survey data. Researchers using even purposefully designed samples frequently have to collapse all non-white ethnic groups into a catch-all ‘Black’ category, which obscures important differences between the ethnic minority groups. Because of their relatively large sample sizes and clear differentiation of ethnic groups, the 1991 SARs permit detailed analysis of the ethnic differences in education, class, employment, living arrangements and other aspects of life chances.

Take education and class for example. Owen et al. (1997) show that Chinese, Black-Africans, Other-Asians, Indians and Other-Others were among the best-educated groups whereas Pakistanis, Bangladeshi, Black-Caribbeans and Black-Others were consistently among the least qualified. This profile obtains whether we look at the male or female groups, or those born in or outside the UK. Ethnic differences in class attainment were also pronounced. Heath and McMahon (1997) show that while 32 percent of British-born White males aged 21–64 were in professional and managerial (service-class) positions, only 10–14 percent of the first-generation Bangladeshi, Black-Caribbean and Pakistani groups were in similar positions. White women were about twice as

likely as Bangladeshi and Pakistani women to be in service-class positions. Moreover, educational qualifications do not translate into occupational positions equally well for all ethnic groups. Among the second-generation and for both sexes alike, each of the ethnic minority groups were significantly less likely to be in service-class positions than British-born Whites, with the sole exception of Black-Others. All this holds true even when age and educational qualifications are taken into account. This, the authors hold, suggests an 'ethnic penalty' for the minority groups.

The research potential of the 2001 SARs in this regard will need to be placed in the context of the changing population, education and class profiles. Recent research has shown that the proportion of ethnic minorities in the UK has sharply increased, from 5 percent in 1991 to around 10 percent in 2001 (Travis, 2001); that education and class profiles have continued to improve since 1991 at the societal level (Li et al., 2002b); and that second-generation ethnic minorities have improved their educational qualifications, employment situations and class positions as against their first-generation counterparts (Dale et al., 2002). Within this context, it would be important to see, in the 2001 SARs, the educational and class attainment of the ethnic minority groups as compared both with the white majority in 2001 and with their situation in 1991. Are ethnic minority groups catching up in education and class, at least for those born and educated in the UK? Do similarly educated people obtain similar social positions? If not, how much of the class differential is due to ethnicity and how much due to other factors? This kind of analysis, together with the religion variables now available, would reveal the relative effects of cultural influences in the interplay between class, education, ethnicity and religion.

The greater differentiation and availability of educational and professional qualifications together with full details of employment status for people aged 16 and over will prompt research of both academic and policy importance. Such research will be highly valuable in terms both of 'the measurement of shortfalls in the provision of education and training and monitoring of take-up of government initiatives', and of 'the assessment of the reserves of skilled and qualified people among the unemployed, including those not currently seeking work' (House of Commons, 1999, paragraph 74, cited in Gardiner, 2002: 42). With regard to educational take-up amongst young people, for instance, one may study the differences between sexes, family origins and ethnic groups in 2001, and compare with the situation in 1991. This kind of analysis would demonstrate whether there is any significant loosening of origin influence over time, or whether class backgrounds still exhibit a powerful resistance to social change. The relevance of the information on professional qualifications is also clear. For example, among people with professional qualifications, who are unemployed or working in jobs 'beneath' their qualifications? Apart from age, gender and nativity factors, what effects would be due to ethnicity? Such an analysis would have clear implications for the implementation of equal opportunity recruitment policies (Dex and Dale, 2001).

The availability of the new class schema (NS-SEC) will be of particular interest to sociologists. Although sociology may not be all about class (Stinchcombe, 1968), it is true to say that no other area in sociology witnesses a more heated and sustained debate than in class analysis. It is difficult here to predict the kind and amount of analysis that will be conducted using the NS-SEC and we can only suggest some broad strands. Is the UK still a class-ridden society or has class receded into history? Frequently, sociologists read an obituary to class analysis (Pakulski and Waters, 1996), but such statements are as often found to be premature as they are made (Goldthorpe, 2000; Goldthorpe and Marshall, 1992). With the 2001 SARs, one could test class differences and compare the situation with the 1991 SARs with respect to a whole range of issues such as housing, health, employment, social deprivation and educational participation of young people.

It is also possible to use the SARs to test competing sociological theories. For instance, although the SC and SEG schemas lack a conceptual basis, one can make theoretically guided comparisons between conceptually based class schemas, such as between the Goldthorpe et al. (1987) and the 'assets' classes (Savage et al., 1992). The two theories debate the nature of social class, especially concerning the class character of professional and managerial employees. The former predicts a growing convergence in professional and managerial class formation whereas the latter foresees the opposite. One may compare their construct validities against a range of issues that are of crucial importance to people's lives, such as long-term unemployment, long-term limiting illness or young people's drop-out from educational participation. Such an analysis, much facilitated by the available categories of higher and lower grades of professionals and managers as shown under the heading of 'Full Version' in Figure 3, would help ascertain the relative explanatory powers of the competing theoretical paradigms (Li, 1997, 2002).

Finally, one may use the range of variables available in the SARs to study the existence and the extent of an 'underclass' in British society. Depending on one's theoretical understanding of what would constitute an underclass, there are a lot of variables from which such a measure could be construed, such as unemployment status and length, education, number of earners in household and living conditions. Furthermore, such an analysis can be conducted at both individual and household levels.

Employment and the Labour Market

It is well acknowledged that labour market participation has an impact on people's lives in terms not only of material rewards but also of mental well-being (Marsh, 1988) and social-capital generation (Li et al., 2002a). The availability of employment status, number of hours worked and other socio-cultural-demographic variables in the SARs permit a variety of analyses of employment situations at the individual, familial and household levels whilst controlling for other factors. For instance, using the 2% SAR, Blackburn et al. (1997) show

striking differences between ethnic groups in their vulnerability to unemployment, even among people with the same level of educational qualifications. Twenty percent of UK-born Black-Africans with higher qualifications were unemployed, but the rate for similarly qualified UK-born Whites was only 3–4 percent. This is another case of ‘ethnic penalty’ (see also Fieldhouse and Gould, 1998).

Another aspect of research is the linking of the SARs with the ONS Longitudinal Study (LS) between 1981 and 1991. Holdsworth and Dale (1997) use the 1% SAR and the LS to explore variations in patterns of employment and occupational attainment among women from different ethnic groups. They show that responses to childbearing are stronger for white women than for most other groups – except Pakistanis and Bangladeshis. Generally, it is only white women who have high levels of part-time working. Whilst those under 35 with no dependent children were unlikely to be economically inactive, the likelihood increases sharply over the other life-stages, especially for white women with unemployed partners or for the unpartnered with dependent children. Among the ethnic minorities, Pakistani-Bangladeshi women were generally found to have the highest odds of economic inactivity across the life-stages. Educational qualifications also played a significant role. Even with life-stages controlled for, Black-Caribbean, Black-Other, Pakistani and Bangladeshi women were about 13 times more likely to be economically inactive than people least likely to be constrained from employment. The study also shows that cultural norms were being modified by socio-economic factors. Although Pakistani-Bangladeshi women were generally disadvantaged, 97 percent of those under 35, UK-born, highly educated, unpartnered and with no dependent children were economically active. Using the LS data, Holdsworth and Dale show that ethnic minority women had higher levels of full-time working over the ages of childbearing. One may think that this would be advantageous to their career progression, but it was not the case. Although more women from minority ethnic groups with children were working full time at both time points, they did not benefit from this but were trapped in poorly paid jobs. Ethnic minority groups were thus doubly disadvantaged: ‘they are more likely than White women to retain a full-time profile and to be in manual jobs’ (1997: 453).

Gender differences in particular occupations proved a fruitful channel for research thanks to the large sample size and type of occupational data available in the SARs. Sociological studies have long noticed the existence of occupational segregation, otherwise called ‘ghettoization’, where men’s and women’s jobs are fairly clearly demarcated. For instance, most senior managerial positions are occupied by men but almost all junior secretarial jobs are done by women (Crompton, 1986). Is there gender inequality in occupations where male and female incumbents have similar characteristics? Survey data usually do not allow detailed analysis for specific occupations because of the sample size but this problem is less serious with the SARs.

As an example, Hakim's study of pharmacy (1998, Ch. 9) is discussed in more detail here because of the straightforward research design and the relatively simple techniques used. Pharmacy is an 'integrated' occupation with roughly equal numbers of male and female occupants who have similar personal characteristics. This is also an occupation where there are exceptionally high proportions of ethnic minority men and women and where there have always been plenty of jobs so that pharmacists could choose their work arrangements to suit themselves. The research questions are: are there gender differences in such an integrated occupation? If so, in what respects, to what extent, and why?

To answer these questions, Hakim used the 1% SAR because of the availability of detailed classifications of occupation and educational qualification subjects which enabled her to identify pharmacists – around 300 were found. The author then used other variables in the file and compared male and female pharmacists in their various profiles. The patterns show that although the two groups were similar in terms of age, marital status, children, ethnicity, education, class and social status, there were considerable differences between them. Male pharmacists were much more likely than their female counterparts to be employers/managers or self-employed, and to work full time and long hours (Hakim, 1998: Tables 9.4–9.7). The findings, the author says, reflect 'fundamental differences in work orientations, entrepreneurship, and risk-taking' (1998: 230). The author concludes that an 'insistence on paid work being fitted around familial responsibilities and a preference for convenience factors over high pay mean that women will generally be concentrated in the lower grades of professional and management occupations, even in the absence of sex discrimination, while men will continue to take the lion's share of higher grade jobs' (1998: 234).

The availability of the 2001 SARs opens up the prospect of comparisons over time. For instance, one could analyse the labour market situation at the individual or the family level and compare the trends over time. The lack of income data in the census is a serious drawback for social scientists, but the data on employment status, hours worked, social class and education can be combined in a way that permits derivation of predicted income.⁸ The availability of data on religion and on professional qualifications in the 2001 SARs will be of particular importance to researchers. In analyses using the 1991 SARs, ethnicity and education are usually taken as surrogates for cultural traits. Lacking religion data, it is not possible to show how ethnicity varies by religion. Where ethnic differences abound, the research findings cannot show 'within-group' differences. With the 2001 SARs, it will be possible to show, in a refined and flexible manner, the interplay of ethnicity, religion, education, class, nativity, life course and gender differences in labour market participation or unemployment risks. Such studies would add significantly to our understanding of the processes and effects of 'acculturation' (Dale et al., 2002). To take a concrete example, research frequently shows that Indians fare very well in British society across a wide range of areas, but Pakistanis and Bangladeshis fare much

worse. Why are there such differences? It would be interesting to see whether the Muslims among the three groups who were all born and educated in the UK have similar attainments in class, education and employment.

Religion

We have already mentioned one possible use of religion in the SARs in analysing its impact on other aspects of life. Another perspective would be to regard it as an outcome variable. For instance, one could investigate the willingness or the lack of it to acknowledge one's religious affiliation as a social barometer. In this regard, Northern Ireland may serve as an example. In the 1991 Census of Northern Ireland, people were asked to 'state the Religion, Religious Denomination or Body ... if none, write NONE'.

Who would write 'NONE' or refuse to provide information, and why? The question is addressed by Macourt (2001) on the basis of the SARs and the SAS (Small Area Statistics) from the 1991 census. He shows that 4 percent of the population 'escaped' the community divide by answering NONE and a further 7 percent by not answering. Further analysis shows that 'middle class people of Catholic community background were very much more likely to answer NONE if they lived in mixed EDs [enumeration districts] than if they lived in exclusively Catholic areas;' and that 'almost all of those from a Catholic background who answered NONE did so living outside the exclusively Catholic areas.' In other words, the Catholics had 'escaped' before they answered NONE. What about the Protestants? The analysis shows that 'very many from a Protestant background who answered NONE did so within exclusively Protestant areas – the "escape" factor was far less marked' (2001: 9). Apart from the area effect, there is also a clear class effect. The higher one's social position, the more likely that one is to 'escape'. Nineteen percent of professionals were escapers, but the figure for the unskilled workers was only 9 percent.

What is the situation in 2001? Is there a growing or decreasing proportion of religious non-affiliates? If class and education profiles continue to improve in Northern Ireland as in the rest of the UK (there is no reason to suppose otherwise) and if socio-educational factors continue to be associated with the non-affiliate behaviour in 2001 as they were in 1991, then one would expect an increase in the proportion of 'nones' and 'no answers' in Northern Ireland in 2001. But would that prediction bear scrutiny? Would there be significant changes in the relative effects of the major determinants of religious affiliation between the 1991 and the 2001 censuses? Moreover, is the 'None' response merely an 'escaping' behaviour or does it represent a genuine disaffiliation? Is it a solely individual behaviour or is there a family effect? Do religious non-affiliates tend to marry each other? Do highly-educated people tend to be more likely to form mixed-religious unions? What effects do people's 'religious backgrounds' play? Who are likely to deny both their religious affiliation and their religious background? Finally, not only can a diachronic study be conducted, a synchronic and cross-national analysis can also be made, such as with the

Republic of Ireland, or with Scotland in terms of both religious affiliation and religious background, or with England and Wales in terms of religious affiliation.

Such questions are not merely of academic interest but of policy implications as well. Whether an increased, similar or decreased proportion of people refuse to acknowledge their religious affiliations, a study of such allegiance would contribute to the development of benchmark data for use by employers and public bodies, particularly the North Ireland Equality Commission, in fulfilling their statutory obligations to avoid discrimination. Research on the housing situation of couples of mixed religions would be of considerable interest and importance to the Northern Ireland Mixed-Marriage Association (NIMMA) which has recently called on such research, and to the Northern Ireland Housing Executive. Finally, a theoretically informed, methodologically rigorous and policy-oriented study on religious affiliation and mobility in Northern Ireland would inform debates on the changing nature of a deeply divided society.

Health and Caring

The data on limiting long-term illness in the 1991 SARs together with all other merits noted earlier have prompted a wealth of analysis into the patterns and causes of morbidity. Research has shown that socio-economic conditions and family circumstances play a major role, especially among the elderly and between different ethnic groups.

The inclusion of the institutional population makes the SARs particularly good for studying morbidity. Glaser et al. (1997) show that previous studies based only on private households have underestimated the prevalence of illness among older people. For example, social surveys suggest better health of single elderly people than elderly married people. But elderly people with poor health are more likely to be living in institutions than in their own home, particularly if they are single. If the institutional population is included, the relationship is reversed. This shows that health needs to be studied from a household, rather than just an individual perspective, and that institutional populations cannot be ignored.

The 2001 SARs will provide more opportunities and challenges. We now have a more ageing population and the proportion of people with limiting long-term illness, which is strongly related to old age, is likely to increase compared with 1991. Apart from this, there are additional health related questions in the 2001 SARs: a general health question and a question on hours spent caring. The combination of these questions will greatly enhance research opportunities. Existing research shows that ethnic minority groups, particularly the South Asians, tend to have large families, and that social deprivation varies between ethnic groups as well as between different socio-economic groups (Dale et al., 2000). Taking these factors into account, we can explore causes of family poverty at a greater depth. For instance, if some ethnic groups are more likely

to live in extended, multi-generational households and to have some members with poor health and limiting long-term illness and other members in lower socio-economic positions, then a disproportionate number of such households would be found in poverty, and their female members may be compelled to spend a lot of time caring for ill family members, resulting in their own economic inactivity and adding to the financial hardship of the household. Religious considerations, cultural norms, financial constraints and language barriers may prevent the households from sending their ill family members into institutional residence. Such families suffer from multiple disadvantages.

Household and Family Composition

The SARs are of great value in understanding household and family composition. The hierarchical structure of the 1% SAR, in particular, allows interrelationships to be explored between individuals within the same household or family, thus facilitating the detailed analysis of family and household composition including inter-ethnic and, for the 2001 Household SAR, inter-religious unions. Research is also available on differences in patterns of leaving the parental home and in household/family formation, particularly with regard to the role of marriage and cohabitation.

In terms of the ethnic mix of family structure, Holdsworth and Dale (1997) show that only a small minority (15 percent) of Black-Other children have both parents of the same ethnicity whereas an overwhelming majority (over 96%) of White, Indian and Chinese children have both parents of the same ethnic group. Over two thirds of Black-Other children (69%) have both parents of different ethnicity. Among Black children of mixed origin, nearly two thirds (65%) have a White mother and Black father, reflecting the highest number of inter-ethnic unions of this type.

With regard to family formation, leaving home and cohabitation among ethnic groups, Heath and Dale (1994) show, for people aged 16 to 35, significant differences amongst Black, White and Asian men and women, particularly with regard to the role of marriage and cohabitation. For example, between ages 22 and 28, Asian men, and to a lesser degree Asian women, were much more likely to live with parents than were White and Black men and women. As compared with Black women, Asian women were also much more likely to be married than cohabiting. Moreover, young Asian men and women with families of their own were more likely to live with their parents or parents-in-law. However, among those born in the UK or with higher educational qualifications, White, Asian and Black women are not much different, although this does not extend to cohabitation.

The family matrix data in the 2001 SARs will increase the accuracy with which family relationships can be defined. As in 1991, it will be possible to analyse not only the presence and the number of dependent children in the family, but also the type of family: whether it is a male- or female-headed family,

whether the partners are married or cohabiting, and how such family types are related to other socio-economic conditions.⁹

Use of the SARs in Northern Ireland, Scotland and Wales, and in an International Perspective

Although social surveys sometimes have ‘boosted’ samples for Wales, Scotland and Northern Ireland, detailed analysis in a particular ‘country’ is often difficult, especially when analyses with refined categories are needed. The SARs help overcome these problems thanks to their large sizes. The following are some examples on Northern Ireland and Wales (for Scottish studies, see Boyle, 1995, 1997).

Research using the Northern Ireland SARs suggests ‘religious penalty’ for Catholics. Borooah (1999, 2000) shows that Catholics were over-represented in the unemployed and under-represented in professional and managerial positions even after age, education and area are controlled for; and that there was evidence of greater social deprivation for Catholics than for Protestants. Is the situation ameliorated in the 2001 SARs? If so, in what respects and to what extent?

The use of ‘country-specific’ variables in the SARs may yield unexpected results. For instance, readers may not be shocked to learn that over 27 percent of ‘Black-Others’ but only 5 percent of Indians were unemployed in Wales in 1991, but may be amazed to find that a particular linguistic ability would give anyone a distinct advantage. Drinkwater and O’Leary (1997) show such a case. Those proficient in Welsh would run much lower risks of unemployment than non-Welsh speakers. Although the two groups were similar in a range of factors such as age or marital status, the former were on the whole better qualified. However, even with these characteristics held constant, the ability to read, write and speak Welsh was found to have a significant effect. Such findings also point to the need to study the role of symbolic and cultural factors, and the impacts of positive discrimination policies in the public sector.

The greater availability of and differentiation in ‘country-specific’ variables in the 2001 SARs can be expected to yield even more vigorous research. For instance, what are the housing conditions and socio-economic circumstances like of Black and Minority Ethnic groups in Wales in 2001? Are their situations better than in 1991 or in England, Scotland and Northern Ireland (Purdam et al., 2002)?

Finally in this section, we give a brief account of research potentials for conducting international comparison using the SARs and their counterparts in other countries (see Dale et al., 2000 for details). Many countries conduct a census and a growing number of them, including the USA, Canada, Australia and Spain, permit census microdata to be sent and used overseas. Even where microdata files are not allowed to go outside the country, comparative research

can be profitably pursued by collaboration between researchers in different countries.

For instance, many studies (Karn, 1997) show that the Chinese in Britain 'do rather well' in comparison with Whites and many other ethnicity groups in a whole range of aspects. Model (2000) uses the SARs, and the US and Canadian counterparts (PUMS and PUMFs respectively) to compare the Chinese socio-economic attainment in the three countries. The evidence shows that the Chinese had greater occupational success in the US than in Britain or Canada. This finding may shed some light on 'American openness' relative to Britain but does not explain why the patterns do not pertain to Canada (see also Erikson and Goldthorpe, 1992). In separate research, Holdsworth (2000) uses the 1% SAR and the US PUMS to compare racial/ethnic family homogeneity in the two countries. The analysis shows far more inter-ethnic unions in Britain than inter-racial unions in New York, suggesting a greater social fluidity in British than American society in conjugality.

Conclusion

The SARs have opened up whole new areas of research that would not have been possible with conventional census (aggregate) or survey data. Not only have the SARs attracted a wide range of users from a wide range of disciplines and sectors, but these users have been prolific in generating high-quality research published in leading academic journals. It is hoped that this article will attract even more users to the SARs to explore the 'gold mine'.

The 2001 SARs present opportunities not only for analysis of social conditions at the turn of the century, but also for comparative research with the 1991 SARs and with census microdata from other industrial and developing countries.¹⁰ The inclusion of new variables such as the new class schema, the greater differentiation and availability of educational and professional qualifications, the religious affiliation and background questions, and the household matrix all provide exciting and unprecedented opportunities given the sample size and rich details of the information contained. We have made some suggestions of possible future analysis which, we would hasten to emphasize, is only a 'drop in the ocean' against the imaginations of the prospective SAR user community.

Acknowledgements

I am grateful to the Editors, the three referees, and Professor Angela Dale and Drs Ed Fieldhouse, Mark Brown and Jo Wathan at Centre for Census and Survey Research (CCSR) of Manchester University for their helpful comments in the earlier versions of the article. I alone am responsible for the errors in the article.

Notes

- 1 For some purposes, researchers are advised to use other data sources. For example, work-life trajectories are better researched using the British Household Panel Surveys, and social values and attitudes are better explored with the British Social Attitude Surveys.
- 2 Owing to the space limit, it is not possible to cover all the SAR-based findings. More information is available at <http://www.ccsr.ac.uk/sars/publications/jointpub.html> and <http://www.ccsr.ac.uk/sars/use/findings/>.
- 3 CCSR has made heavy use of the 1% SAR in teaching hierarchical data analysis. The two popular courses are 'Analysing Hierarchical Surveys' for SPSS users and 'Data Management with Stata' for Stata users. More information is available at <http://www.ccsr.ac.uk/courses/internal/2003-2004/info.html>.
- 4 For 2001 census forms, see <http://www.statistics.gov.uk/census2001/censusform.asp>.
- 5 For specifications of the 2001 SARs, see <http://www.ccsr.ac.uk/sars/2001/request/>.
- 6 Full details are given in <http://www.ccsr.ac.uk/sars/>.
- 7 Full details are available at <http://www.ccsr.ac.uk/courses/>.
- 8 This information is already available in the 1991 SARs and will be added to the 2001 SARs.
- 9 The following example is Stata syntax for constructing the variable `mfhead` (male or female headed family) and how it varies with the number of cars.

```

gen          mhead=1 if sex==1&famhead==1
replace     mhead=2 if sex==2&famhead==1
replace     mhead=0 if mhead==.
egen       mfhead=max(mhead),by(hnum famnum)
keep if    mfhead!=0
bys hnum famnum: keep if _n==1
lab var    mfhead "Male or female headed?"
lab def    mfhead 1 "Male-headed" 2 "Female-headed"
lab val    mfhead mfhead
tab mfhead cars,r

```

	Male or female headed?				Total
	0	1	2	3+	
Male-headed	20372 16.34	61475 49.30	35320 28.32	7531 6.04	124698 100.00
Female-headed	11716 41.92	10937 39.13	4472 16.00	825 2.95	27950 100.00
Total	32088 21.02	72412 47.44	39792 26.07	8356 5.47	152648 100.00

- 10 For instance, collaboration is under way between CCSR and China's Shanghai Social Science Academy on comparative research on British and Chinese social structures using the microdata from the 2001 censuses of the two countries.

References

- Blackburn, R.M., A. Dale and J. Jarman (1997) 'Ethnic Differences in Attainment in Education, Occupation and Lifestyle', in V. Karn (ed.) *Ethnicity in the 1991 Census, Volume Four*. London: HMSO.
- Borooah, V. (1999) 'Is There a Penalty to Being a Catholic in Northern Ireland? An Econometric Analysis of the Relationship between Religious Belief and Occupational Success', *European Journal of Political Economy* 15: 163–92.
- Borooah, V. (2000) 'Targeting Social Need: Why are Deprivation Levels in Northern Ireland higher for Catholics than for Protestants?', *Journal of Social Policy* 29: 281–301.
- Boyle, P. (1995) 'Modelling Population Movement into the Scottish Highlands and Islands from the Remainder of Britain, 1990–91', *Scottish Geographical Magazine* 111: 5–12.
- Boyle, P. (1997) 'Contrasting English and Scottish Residents in the Scottish Highlands and Islands', *Scottish Geographical Magazine* 113: 98–104.
- Crompton, R. (1986) 'Women and the Service Class', in R. Crompton and M. Mann (eds) *Gender and Stratification*. Cambridge: Polity.
- Dale, A., E. Fieldhouse and C. Holdsworth (2000) *Analysing Census Microdata*. London: Arnold.
- Dale, A., E. Fieldhouse, N. Shaheen and V. Kalra (2002) 'The Labour Market Prospects for Pakistani and Bangladesh Women', *Work, Employment and Society* 16(1): 5–25.
- Dex, S. and A. Dale (2001) 'Implementing Equal Opportunities Recruitment Policies', Research Project funded by Joseph Rowntree Foundation.
- Drinkwater, S. and O. O'Leary (1997) 'Unemployment in Wales: Does Language Matter?', *Regional Studies* 31: 583–91.
- Erikson, R. and J.H. Goldthorpe (1992) *The Constant Flux*. Oxford: Clarendon Press.
- Evans, G. and C. Mills (2000) 'In Search of the Wage-labour/Service Contract: New Evidence on the Validity of the Goldthorpe Class Schema', *British Journal of Sociology* 51: 641–61.
- Fieldhouse, E. and M. Gould (1998) 'Ethnic Minority Unemployment and Local Labour Market Conditions in Great Britain', *Environment and Planning A* 30: 833–53.
- Gardiner, C. (2002) 'Improving Information for Regional Development Agencies' Strategies and Policies: Using Microdata from the United Kingdom's 2001 Census of Population', *Environment and Planning C: Government and Policy* 20: 39–55.
- Glaser, K., M. Murphy and E. Grundy (1997) 'Limiting Long-term Illness and Household Structure among People Aged 45 and Over, Great Britain 1991', *Ageing and Society* 17: 3–19.
- Goldthorpe, J.H. (2000) *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.
- Goldthorpe, J.H., with C. Llewellyn and C. Payne (1987) *Social Mobility and Class Structure in Modern Britain*. Oxford: Clarendon Press.
- Goldthorpe, J.H. and G. Marshall (1992) 'The Promising Future of Class Analysis: A Response to Recent Critiques', *Sociology* 26(3): 381–400.

- Hakim, C. (1998) *Social Change and Innovation in the Labour Market: Evidence from the Census SARs on Occupational Segregation and Labour Mobility, Part-time Work and Students' Jobs, Homework and Self-Employment*. Oxford: Oxford University Press.
- Heath, A. and D. McMahon (1997) 'Education and Occupational Attainments: The Impact of Ethnic Origins', in V. Karn (ed.) *Ethnicity in the 1991 Census: Employment, Education and Housing among the Ethnic Minority Populations of Britain*. London: HMSO.
- Heath, S. and A. Dale (1994) 'Household and Family Formation in GB: The Ethnic Dimension', *Population Trends* 77: 5–13.
- Holdsworth, C. (2000) 'Ethnic Composition of Families in Great Britain and the U.S.A.', in A. Dale, E. Fieldhouse and C. Holdsworth *Analyzing Census Microdata*, pp: 200–5. London: Arnold.
- Holdsworth, C. and A. Dale (1997) 'Ethnic Group Difference in Women's Employment', *Work, Employment and Society* 11: 435–57.
- Karn, V. (ed.) (1997) *Ethnicity in the 1991 Census, Volume Four: Employment, Education and Housing Among the Ethnic Minority Populations of Britain*. London: HMSO.
- Li, Y. (1997) *The Service Class: Theoretical Debate and Sociological Value*. D.Phil Thesis, Nuffield College, Oxford University.
- Li, Y. (2002) 'Falling off the Ladder? Professional and Managerial Careers', *European Sociological Review* 18(3): 253–70.
- Li, Y., A. Dale and E. Fieldhouse (2001) 'Key Findings from the 1991 SARs', URL (consulted Feb. 2004): <http://www.ccsr.ac.uk/sars/use/findings/KeyFindings.pdf>
- Li, Y., M. Savage and A. Pickles (2002a) 'Social Inequalities and Associational Involvement in England and Wales (1972–1999)', *British Journal of Sociology* 54(4): 497–526.
- Li, Y., M. Savage and A. Pickles (2002b) 'Social Mobility, Social Capital and Civic Participation in England and Wales (1972–1998)', mimeo, University of Manchester.
- Macourt, M. (2001) 'The Escapers: The "Nones" in Northern Ireland Religion', URL (consulted Feb. 2004): <http://www.ccsr.ac.uk/sars/publications/newsletters/17/>
- Marsh, C. (1988) *Exploring Data: An Introduction to Data Analysis for Social Scientists*. Cambridge: Polity in association with Basil Blackwell.
- Marsh, C., C. Skinner, S. Arber, B. Penhale, S. Openshaw, J. Hobcraft, D. Lievesley and N. Walford (1991) 'The Case for Samples of Anonymised Records from the 1991 Census', *Journal of the Royal Statistical Society (A)* 154(2): 305–40.
- Marshall, G., H. Newby, D. Rose and C. Vogler (1988) *Social Class in Modern Britain*. London: Hutchinson.
- Model, S. (2000) 'A Cross-national Look at Chinese Immigrant Occupations', in A. Dale, E. Fieldhouse and C. Holdsworth (2000) *Analyzing Census Microdata*, pp. 195–200. London: Arnold.
- Office of Population Censuses and Surveys, General Register Office for Scotland (1992) *1991 Census Definitions*. London: HMSO.
- Owen, C., P. Mortimore and A. Phoenix (1997) 'Higher Education Qualifications', in V. Karn (ed.) *Ethnicity in the 1991 Census, Volume Four*. London: HMSO.
- Pakulski, J. and M. Waters (1996) *The Death of Class*. London: Sage.

- Purdam, K. and Y. Li, with M. Brown and J. Wathan (2002) 'A Profile of the Housing and Socio-economic Circumstances of Black and Minority Ethnic People in Wales.' Project report funded by and submitted to the Welsh National Assembly.
- Rose, D. and K. O'Reilly (eds) (1997) *Constructing Classes, Towards a New Social Classification for the UK*. Swindon: ESRC/ONS.
- Rose, D. and K. O'Reilly (1998) *Final Report of the ESRC Review of Government Social Classifications*. Swindon: ESRC/ONS.
- Rose, D. and D. Pevalin (2001) *The National Statistics Socio-economic Classification: Unifying Official and Sociological Approaches to the Conceptualisation and Measurement of Social Class*. ISER Working Papers, 2001-4, Colchester: University of Essex.
- Savage, M., J. Barlow, P. Dickens and T. Fielding (1992) *Property, Bureaucracy and Culture: Middle-class Formation in Contemporary Britain*. London: Routledge.
- Stinchcombe, A. (1968) *Constructing Social Theories*. New York: Harcourt, Brace and World.
- Travis, A. (2001) 'Ethnic Minorities Grow to 1 in 10', *The Guardian*, 23 February 2001.

Yaojun Li

Is a Lecturer in Sociology and Social Statistics, and Director of MA Social Research Programmes at the Department of Sociology and the School of Social Sciences, Birmingham University. His research interests are in social mobility and social stratification, socio-cultural capital, socio-political identity, occupational and educational attainment, poverty and social deprivation, cross-national comparison of class and gender differences, and statistical analysis and modelling of large-scale and complex social surveys. Recent peer-reviewed publications have appeared in *European Sociological Review*, *British Journal of Sociology*, *Work, Employment and Society*, *Ageing and Society*, *Sociology Research Online*, *Sociological Review* and *Explorations in Sociology*. He also has numerous book chapters on research methods, ethno-religious divisions and changing social structures in Britain, which will be forthcoming in 2004/2005.

Address: Department of Sociology, University of Birmingham, 32 Pritchatts Road, Edgbaston, Birmingham B15 2TT, UK.

E-mail: y.li.3@bham.ac.uk