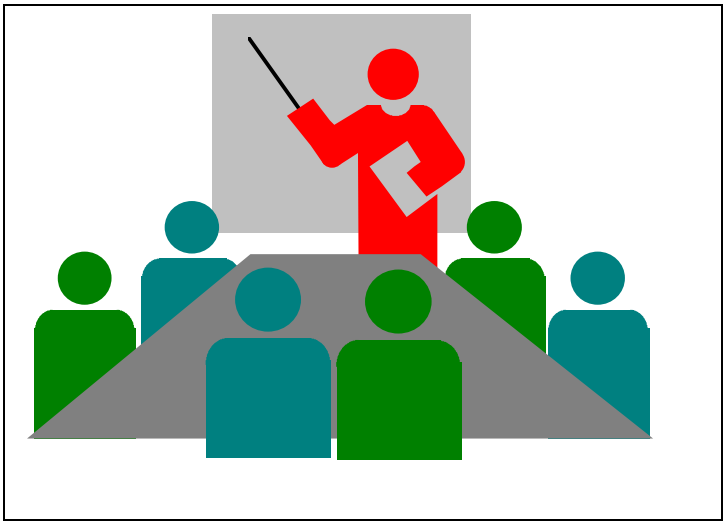


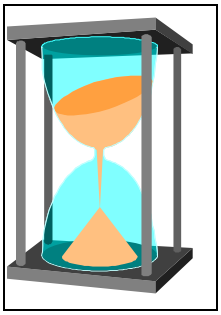
UPDATE -

News from the LS User Group

Issue no. 12
November 1995

Contents: **Page**

This newsletter is designed to provide rapid dissemination of information on the OPCS Longitudinal Study (LS) and a forum for the exchange of users' views and comments. It is produced by the LS User Support Programme at the Social Statistics Research Unit (SSRU) at City University. All comments and con tributions for the newsletter should be sent to Rosemary Creeser, LS Support Programme, SSRU, City University, North ampton Square, London EC1V 0AR tel. 0171 477 8487 Email: rc@ssru.city.ac.uk Contributions on IBM -formatted floppy disk are always welcome and should be sent, clearly documented (file name, wordprocessing package and version used) along with a hard copy of the text.



1 Diary

This section highlights forthcoming events of interest to LS users.

If you are arranging an event and wish to publicise it in future issues of *Update* you should send details to Dina Maher, the LS Administrative Secretary at SSRU.

LS Workshops

SSRU hold regular 2-day workshops to provide detailed information on the study and to enable researchers to gain practical hands-on experience of accessing the data. The workshops also provide an ideal opportunity to meet members of the LS Support Team and to discuss the suitability of the LS for investigating specific research questions. The next LS Workshop will be held on 12th/13th December. A programme and booking form are enclosed with this newsletter.

As part of the hands-on element of the workshop participants are able to specify a statistical analysis of their choice using a small sub-set of variables extracted using the LS FLEXTRACT facility and a test data-set based on 1% of the LS data-base. As usual, the number of places is limited to ensure that participants get sufficient individual attention and hands-on experience. A fee of £20 will be charged to cover documentation and administrative costs. Researchers who plan to start work shortly on projects using LS data, are advised to contact Dina Maher immediately on 0171 477 8486 to reserve a place.

2 Funding for the LS User Support Programme

Funding for the continuation of the LS User Support Programme over the period 1996-2001 has been agreed by the ESRC and the team at City University is due to become one of the Census Units funded under Phase II of the ESRC-JISC Census Programme. Consequently, some of our procedures may have to be adapted to meet available resources. Details of such changes will be announced in future issues of *Update*. We would like to thank all of the users who wrote to us to register their interest in the study.

3 Report on the 1-day LS conference "The enlargement of a complex dataset: some first analyses"

On Wednesday October 18th, a one-day LS conference, organised jointly by SSRU and the ESRC ALCD (Analysis of Large and Complex Datasets) programme took place at City University. ALCD is an ESRC-funded initiative designed to raise the standards of research in UK social science. The programme originated in response to the findings of a report undertaken in the late 1980s, which identified both the need for longitudinal data and the development of methodological work. It incorporates several different components including the provision of support for training and research fellowships/studentships. In the following pages we summarise some of the key points made during the day.

Data quality: linkage and coverage, Lin Hattersley, OPCS

The LS was set up in the early 1970s to be a representative 1% sample of the enumerated population of England and Wales. The population has been re-sampled at each census since 1971. Population change is reflected by the intercensal addition of new members, born on LS dates, and the recording of those exiting the study either by death or emigration. Members are linked between censuses and the events occurring to them (live-births, still-births, cancers etc) are also linked into the study.

Two major questions may be considered in relation to the quality of the data. How representative is it and how good is the coverage, both of the base population from which the sample is drawn at census, and the events that are subsequently linked to that population?

As the overall sampling fractions in figure 1 show, the LS represents approximately 1% of the enumerated population at each of the three censuses. (Figure 1 is not available in the .PDF version of *Update*.) A comparison of the age and sex distributions of each of the census-LS samples with those for the England and Wales population shows that, with the exception of two small sub-groups, sampling variation is extremely small with no obvious over or under-sampling. The sub-groups in question are those born outside the UK, who are known to be over-sampled, and those living in communal establishments who are known to be enumerated with poor data on birth dates.

At each census a link is made between the LS sample present at the previous census and those in the current census-LS sample. 91% of LS members were successfully linked between the 1971 and 1981 Censuses and 90% between the 1981 and 1991 Censuses. The best linkage rates are found for those LS members who were middle-aged in 1981 or approaching retirement in 1991, living in stable social conditions and in employment (see table 1). As might be expected, the worst linkage rates are found for the younger, more mobile sub-groups of the LS sample. A major cause for linkage failures appears to be the inconsistent reporting of birthdates from census to census.

Events such as births, deaths, cancer registrations etc, are linked to LS members using the National Health Service Central Register. The linkage rates for events are calculated as a ratio of observed to expected events occurring to the LS population. Most of these linkage rates are extremely good - around 100% for both new births into the sample and deaths of sample members, and over 90% for births to female LS members. However, the linkage rates for migration data (immigrants joining the sample as new members and emigrations of LS members out of England and Wales) are poor as a result of the numerator/denominator mismatch.

On the whole, considering the size of the sample (in excess of 820,000) and the complexity of the data, the quality of the LS is extremely good. Coverage of 1% of the population is achieved and linkage rates for both census and event data are high. While some inconsistencies have been found these are comparatively few .

Table 1: A comparison of the best and worse linkage rates

1971-81 Census-LS link	Linkage rate	1981-91 Census-LS link	Linkage rate %
Best linkage rates	%		
1971 characteristics:		**1981 characteristics:**	
Aged 35-39	93	Aged 50-64	93
Married	92	Married	92
Living in private household	92	Living in private household	90
Owner-occupier/LA tenant	92	Owner occupier	92
Household of 3-6 persons	92	Household of 2 persons	92
In employment	92	In employment or other inactive*	92
Social Class IIIN	93	Social Class IIIN	93
Born in England and Wales	92	Born in England and Wales	92
Worst linkage rates			
Aged 15-19	89	Aged 10-19	85
Divorced	88	Single	87
In communal establishment	79	In communal establishment	76
Renting furnished housing	79	Renting furnished housing	76
Household of 7+ persons	87	Household of 7+ persons	81
Unemployed	86	Student	81
Social Class - Armed Forces	81	Social Class - Armed Forces	81
Born in Caribbean/NCW	69	Born in Far East/NCW	75

*Key: * includes housewives and those of independent means. NCW=New Commonwealth*

Some consistency checks on the reported ages of the elderly in the LS, John Craig, OPCS

The accuracy of the reporting of the age of the elderly is of particular interest to users of the census, such as the NHS and local authorities, who depend upon the calculation of the percentage of elderly aged 85+ in a given area.

Two consistency checks on the age of the elderly have been carried out using the LS. The first compared the age given at the 1991 Census with the age of the same individual from the 1971 Census. The second focused on those LS members who died in the twenty year period following the 1971 Census, and compared the age from the 1971 Census with that implied by the information provided at death registration.

The provisional results of both checks show that there is some inconsistency in the dates of birth, and hence age. However, the amount is small - at most ages about half of the inconsistencies are of one year. As figure 2 shows, inconsistencies increase with age, especially for those aged over 70 in 1991, and are greater for women than for men. (Figure 2 is not available in the .PDF version of *Update*.) One possible explanation for this is that in 1971 women, particularly older women, were considerably less likely to have completed the census schedule than men. Figure 2 also shows that age inconsistencies are asymmetrical and do not wholly cancel out with those in the other direction. A detailed summary of these findings will be published in early 1996.

Inter-regional migration and intra-generational social class mobility 1971-91[1] Tony Fielding, University of Sussex

This presentation summarised some of the first results to use data from the 1971, 1981 and 1991 Censuses now available in the LS. The paper explored the following issues:

i. Is there a relationship between inter-generational social mobility and inter-regional geographical mobility ? Are those who are more geographically mobile also likely to be more socially mobile ?

ii. How are the middle classes formed ? For example, in 1981 what were the geographical and social class locations of those who were members of the three middle classes (professionals, managers and 'petty bourgeoisie') in 1991 ?

iii. Once formed, how geographically mobile are the members of the three middle classes ? Are professionals more or less mobile than managers ? How does middle class spatial mobility compare with working class spatial mobility ?

iv. How stable are middle class memberships ? Where, both socially and geographically, do people go when they cease to be professionals, managers and members of the "petty bourgeoisie" ?

v. Are certain locations more important than others in middle class formation ? Specifically, does the South East Region play a special role in the construction of middle class careers ?

vi. Do men and women face similar "geographies of opportunity" or are there gender differences in the spatial patterns of social mobility ? If the patterns are different, how does this affect the benefits from migration for men and women ?

The main conclusions of the paper are that social mobility and geographical mobility are positively related. Those who change class locations are also likely to change their geographical locations.

During the period 1981-91 the South East region attracted a large number of potentially upwardly mobile young people, who were entering the market for the first time. Along with long-term residents of the region, they were promoted socially, at rates which were distinctly above those for the country as a whole. At the same time, as a result of out-migration, the region lost a significant proportion of those who had been promoted, who were at a later stage of their working lives or approaching retirement. The region is therefore said to have acted as an (upward) social class "escalator region" in 1980s, as it had also done in the 1970s.

It appears that while migration to the South East implied upward mobility for both men and women, it was particularly favourable for the development of women's careers. However, out-migration from the region implied sideways mobility for men but downward social mobility or the termination of a career for women.

Finally, the paper concludes that class-specific inter-regional migration rates changed very little between the 1970s and the 1980s.

---

[1] A version of this paper was published as: Fielding A J (1995) "Inter-regional migration and intra-generational social class mobility 1971-1991", in Savage M and Butler T (1995) (Eds), Social Change and the Middle Classes, London:UCL

Multi-level modelling: some preliminary findings on the relationship between limiting long-term illness, social mobility and work history, Simon Gleave and Dick Wiggins, SSRU

Simon Gleave and Dick Wiggins[2] presented some of the preliminary findings of a multi-level model of the relationship between limiting long-term illness and a sub-set of variables summarising an individual's labour market history. For the purposes of this presentation they focused on a group of men aged 16-45 who were enumerated at the 1991 Census in Wales.

The level 1 explanatory variables included those measuring the individual's educational level, whether they had changed social class between 1971 and 1981 (eg 'moved up', 'moved down', 'stayed static') and had ever lived in the South East. Only one variable was included at level 2 - an indicator of the district in which the men were enumerated.

The final model is shown below with the significant estimates. However it is the variance terms that are mainly of interest. The variance between districts is very small which suggests that the incidence of limiting long-term illness is fairly stable across the 37 Welsh districts considered in the analysis. The variance between individuals has an age factor to it and further study may reveal more complex level 1 variance.

It is hoped to extend work on this project to look at the effect of adding a quadratic age term and to examine the whole of England and Wales.

Table 2: Results for the final model

Parameter	Estimate	Standard Error
Constant	-0.06501	0.06034
Age	0.00548	0.00231
No degree	-0.13620	0.06341
Inactive	-0.11250	0.05089
Age*inactive	0.01129	0.00186
Age*no degree	0.00783	0.00243
Variance between districts	0.001529	0.00063
Variance within districts:		
Constant	-0.02255	0.00432
Age	-0.00013	0.000006
Likelihood	3370.94	

Further multi-level modelling using the LS

This particular analysis concentrated on those enumerated in Wales as the multi-level software (ML3) installed on the OPCS computer would not initially work with the full LS dataset. Future analyses should not be restricted in this way as the technical problem has now been resolved.

Researchers keen to carry out multi-level analyses may be interested to learn that it should soon be possible to specify a three or four level structure. Furthermore, in the future the specification of the model may be enhanced by incorporating compositional variables such as the 1991 Small Area Statistics (SAS) at the second level.

[2] Dr Dick Wiggins is Director of Graduate Studies at SSRU. He is currently taking a six month sabbatical leave, funded by the ESRC ALCD programme, to carry out methodological research with colleagues at the Institute of Education .

Ethnic variation in occupational consequences of motherhood, Clare Holdsworth, Cathie Marsh Centre for Census and Survey Research, University of Manchester

It is well-established that women tend to leave the labour force on childbearing and return to employment, which is often part-time, once their children are grown up. Though this trend has been confounded in recent years with more women choosing to retain their jobs and take maternity leave, the overall pattern of women's employment participation associated with childbearing tends to lead to downward occupational mobility and a loss of lifetime earnings. However, this pattern is by no means universal across all groups and recent analyses of women's employment participation by ethnic group have highlighted some important differences. Generally, a much larger proportion of White women are found in part-time employment than their Black or Asian counterparts. Furthermore, by comparison with White women, far more Black women continue in full-time employment after the birth of their children. Pakistani and Bangladeshi women tend to leave employment altogether, although Indian women continue to work, usually in full-time jobs.

This presentation summarises work which is currently underway, using both the LS and the 1991 Census SARs (Samples of Anonymised Records), to explore the ethnic differences in this relationship.

Cross-sectional analysis

To date, the 1% Household File from the 1991 SARs has been used to develop a cross-sectional model of the relationship between ethnic group and 1991 employment status. For this component of the work a "continuation ratio model" has been used - the results of which are presented in terms of the odds ratios found in table 3. This particular type of model divides the dependent variable, employment status, into two partitions. The first compares the relative odds of woman working against not working for each of the explanatory variables, while the second compares for all working women the odds of working full or part-time.

Focusing on the results for the first partition, we see that women with an unemployed partner were significantly more likely to be "not working" than those with a partner in work. Although the relationship is less strong the results for the second partition ('part-time versus full-time') show a similar pattern.

One of the most striking results is the increased propensity of women with young children, particularly those heading one-parent families, to be working, compared with the baseline category of young single women. For working women, women with children are more likely to be working part-time rather than full-time. The table also highlights that, compared with the baseline category of White women, those in the major ethnic groups were significantly more likely to be working full-time than part-time in 1991.

Table 3: Results of modelling the 1% Household SAR - Continuation Ratio Model

	1st partition			2nd partition
	Not working V Working			PT V FT
Variable & base category	**Parameter Estimate**	**Standard Error**	**Odds Ratio**	**Odds Ratio**
Lifestage: No partner < 35, ndpc				
No partner > = 35, ndpc	1.25	0.03	3.5	3.9
Partner < 35, ndpc	-0.48	0.03	0.6	1.1*
Partner, dpc < = 4	2.07	0.02	8.0	28.7
Partner, dpc 5-9	1.06	0.03	2.9	29.7
Partner, dpc 10+	0.68	0.03	2.0	15.5
Partner > = 35, ndpc	0.79	0.02	2.2	9.8
No partner, dpc < = 4	3.00	0.04	20.2	13.6
No partner, dpc 5+	1.61	0.03	5.0	11.3
Partner's emp status: In work				
Partner out of work	1.26	0.02	3.5	1.2
Higher qualifications: None				
Below degree qualifications	-0.96	0.03	0.4	0.6
Degree plus	-1.00	0.03	0.4	0.4
Ethnic group: White				
Black Caribbean & Other	-0.21	0.05	0.8	0.4
Black African	0.50	0.10	1.7	0.5
Indian	0.21	0.05	1.2	0.3
Pakistani & Bangladeshi	2.00	0.08	7.4	0.6
Chinese	0.03	0.11	1.0*	0.6
Other Asian	0.37	0.09	1.4	0.5
Place of birth: UK born				
Not UK born	0.22	0.26	1.2	0.9

*Key: ndpc=no dependent children, dpc=dependent child, * =Not significant at the 5% level.*

Longitudinal analysis

The next component of this research will use LS data to analyse the change in women's economic activity and occupational attainment over the period 1981-91. Ideally, it would be useful to have access to longitudinal work history data to establish the precise ordering of changes in occupation in relation to childbearing. By comparison with other large-scale British longitudinal studies, such as the British Household Panel Study (BHPS) the occupational data in the LS is mainly restricted to that collected at the 1971, 1981 or 1991 Censuses. (The recording of occupation at vital events such as the birth of a child is less complete.) However, one of the reasons for using the LS is the availability of the Cambridge Occupational Score(s), which allow a finer distinction to be made between occupational attainment at each of the census points. We hope to include some of the preliminary results of this work in future issues of this newsletter.

Indicators of deprivation in longitudinal perspective, Andy Sloggett, Centre for Population Studies, London School of Hygiene and Tropical Medicine

Andy Sloggett presented an account of work in progress with Heather Joshi which is using the LS to unpack the association of social deprivation, often measured at an area level, and individual experience of living conditions and life events.

The presentation set the context for current work by reporting associations of neighbourhood deprivation, measured at ward level through Small Area Statistics, with various life events in the LS, all clearly or potentially adverse. The local deprivation score is a combination of four variables: low social class, rented housing tenure, lack of car access and unemployment. The outcomes illustrated were premature mortality, low birthweight, stillbirth, teenage motherhood and solely registered births outside marriage. Once information about individual circumstances was included in these models, it appeared that most, if not all, of the high incidence of adverse outcomes in the more deprived areas could be explained by the particular circumstances of the individuals. The one exception was still births which showed little systematic social variation. This suggests that area indicators should be used as indicators of the location of disadvantaged people, but casts some doubt on the idea that these indicators identify independent area effects exacerbating the disadvantages of people wherever they live.

The rest of the presentation reported on further exploration of the experience, by individuals, of the four conditions used to identify deprived areas (plus permanent sickness as another response to the census economic activity question). The cross-sectional results summarised in table 4 show that the rate of reporting either unemployment or permanent sickness in the cohort aged 16-49 in 1971 was generally quite low at each of the three censuses (eg less than 8%). In contrast , the proportion in Social Classes 4 or 5, without access to a car or living in rented accommodation were each at least 16%, rising to 48% for tenants in 1971. Relatively few experienced multiple deprivation. Counting either of the mutually exclusive unemployment or permanent sickness as 1, those with 3 or more deprived states were between 9% and 5% of the cohort considered.

Looking longitudinally at the cohort's experience of single and combined states across three censuses, a fair amount of turnover was observed. Those in a given state at any one time were unlikely to have been in it at the other censuses (permanent sickness excepted). Also the experience of the state at **any** census extends beyond those people picked up at a cross section. Despite quite broad bands between those experiencing the state at **all** three censuses and the number who have been included at **any,** the range is not generally as great as would be expected if rotation in these states was completely random. There are indications of persistence, or serial correlation, but it is not complete. Just as people move in and out of deprived areas, they move in and out of deprived states.

Table 4: The proportion in 5 "deprived" states at the 1971, 1981 and 1991 Censuses#

	Women			**Men**		
	1971	**1981**	**1991**	**1971**	**1981**	**1991**
Permanently sick	0.4	1.1	4.4	0.3	1.5	8
Unemployed	2.2	2.9	2.2	3.6	7.5	6.4
Social Class 4 or 5*	16	19	21	20	19	18
No car access	34	23	22	32	19	16
Rented accommodation	48	34	23	48	33	21
3+ deprived states	6	5	5	9	8	7
# Sample: 86,624 women and 83,771 men aged 16-49 in 1971, present at each census.						
* current or last reported occupation of LS member was unskilled or semi-skilled.						

Logistic regression was used to model the chances of being multiply deprived in 1991, conditional on the number of states reported in the previous census. Multiple disadvantage in 1991 was strongly associated with past disadvantage. For example, focusing on the number of states in 1971, males who had been clear of any deprivation had odds one fiftieth smaller that those with 3 states, of experiencing three or more states in 1991. For females the odds ratio was 1 in 30. An alternative model of the 1981 deprivation states gave an improved fit, with the gradient steepening between those with increasing numbers of deprivations. However, the results of models based on information from both 1971 and 1981 indicate that the 1971 state is not redundant. Those who had escaped from deprivation in 1981 had not necessarily escaped for ever.

Preliminary models of a further adverse outcome, limiting long-term illness in 1991, were also presented for females. This too was related to the number of deprivations at both previous censuses, more strongly to 81 than 71. Taking the individual components of the deprivation index in 1981, we find, not surprisingly, that permanent sickness in 1981 was by far the most strongly associated with illness in 1991 (Odds Ratio=49). After that the next closest association was for unemployment (OR=2), then no car (OR= 1.7), and tenancy (OR=1.6). Controlling for the other factors, membership of social classes 4 or 5 was associated with an odds ratio 1.1 times those of the non-deprived reference group. As predictors of multiple deprivation rather than long-term illness, tenancy, no car and low social class were more strongly associated than the smaller categories of sickness and unemployment.

The next stage of the research is to integrate the histories of individual variables with the deprivation scores associated with the wards in which the individuals were enumerated.

The research projects reported by Clare Holdsworth and Andy Sloggett are both supported by the ESRC Census Programme.

4 Technical issues

Using the LS for intra household analyses, Judith Wright and Kevin Lynch

The LS is a sample of individuals, but it is also a valuable source of information about the household in which the LS member (LSM) is enumerated at census. It allows us to explore both the type of family in which the LSM is living and the household composition. Variables such as those denoting the "family type" and the "minimal household unit" facilitate this type of analysis.

In addition, it is possible to investigate particular relationships *within* a family or household - for example, a parent of an LSM, or their son, daughter or spouse. This allows researchers to identify whether there has been an inter-censal change in the individuals living in the same household as the LSM: a change in spouse or parent. It is also possible to identify specific relationships which allow us to assess the prevalence of particular circumstances - such as cohabitation. For example, what proportion of (LSM) children in lone parent families living with "other adults" present are related to these other adults ?

The following comments cover some of the technical issues involved when using the LS for this type of research. The next issue of *Update* will explore the separate issue of how to confirm whether the same person has been identified at subsequent censuses.

Identifying relationships in the household

The source of information about relationships in the household is derived from the census question *relationship in household,* which is asked of all people in the household with reference to the first person on the census schedule (the 'head of household'). While some LS relationship variables are the product of census processing and tabulation, others have been specially derived for researchers using the LS. These variables are described in section 8.5 of the *LS User Manual* and *LS User Guide No 1, Households, families and fertility,* (Penhale, 1990).

For 1991, a similar distinction is made between variables produced from census processing and those which have been specially derived. It is from census coding of *relationship in household* (RELAT) that the *relationship to the LSM* (LSRELAT) variable is derived. The following example shows the result of this coding, where the LSM is a child within a married couple family, in a four-person household.

In this case it would be simple to identify the mother of the LSM by searching the non-LS member file for the person in the household with LSRELAT code=4. (The algorithm for LSRELAT uses the sex variable to identify whether it is the mother or father and the RELAT code 01= husband or wife where the LSM is the son or daughter.)

Example 1: The LSM is in the head of household's family

1991 Census Schedule:	RELAT	LS RELAT	FANUM
Person no 1: Head	00=Head	07=Parent	1
Person no 2: Wife	01=Husband or wife	07=Parent	1
Person no 3: Daughter	03=Son or daughter	09=Brother/sister	1
Person no 4: Son	04=Son or daughter	LSM	1

This simple example illustrates how to search for relationships within a household using LSRELAT, where there is only one family in the household. However, more complex household compositions also need to be considered. The 1991 Census variable, LSRELAT, reflects this complexity (the LS data dictionary lists the codes 01 to 43, and some of the codes reflect ambiguous relationships: code 26 = 'Parent or other unrelated'; code 41 = 'Parent, uncle or aunt, or other unrelated').

The complexity arises when the relationship to the LSM is not obvious from the household member's relationship to the head of household. This is the case when the head of the LSM's family is not the head of the household - the LSM is in fact in a "concealed" family - and the relationships in the LSMs family may not be easily identified without using additional variables to LSRELAT. Consider the following examples (*members of the LSM's family are in italics)* :

Example 2: The head of the LSM's family is not related to the head of household.

1991 Census Schedule: (Relationship to head)	RELAT	LSRELAT	FANUM
Person no 1: Head	00=Head	16=Other unrelated	1
Person no 2: Son	03=Son or daughter	16=Other unrelated	1
Person no 3: Daughter-in-law	05=Son or daughter in law	16=Other unrelated	1
Person no 4: Unrelated	*16=Other unrelated*	*16=Other unrelated*	*2*
Person no 5: Unrelated	*16=Other unrelated*	*LSM*	2

Example 3: The LSM is related to the head of household but the LSM is not in the same family as his parents (the LSM is married and in a family of their own).

1991 Census Schedule: (Relationship to head)	RELAT	LSRELAT	FANUM
Person no 1: Head	00=Head	18=Grandparent	0
Person no 2: Son	03=Son or daughter	39=Parent, uncle or aunt	1
Person no 3: Daughter-in-law	05=Son or daughter in law	39=Parent, uncle or aunt	1
Person no 4: Grandson	*11=Grandchild*	*LSM*	*2*
Person no 5: Granddaughter-in-law	*13=Other related*	*13=Other related*	*2*
Person no 6: Great grandson	*13=Other related*	*13=Other related*	*2*

In both examples, it is impossible to identify the LSM's parents in the household by only using LSRELAT.

In the first example (example 2) the LSM is a member of a lone-parent family - the parent being person number 4 on the census schedule. In this case it would be necessary to also use the "generation in family" variable (GENINFAM) to identify the person in the first generation with the same family number (FANUM) as that of the LSM in order to identify the parent.

Use of the GENINFAM and FANUM variables takes advantage of the census definition of a family which means there can only be one or two people in "generation one" in the same family as the LSM.

In the second example (example 3) where the head of the household is related to the head of the family, but the LSM is not in the same family, family number cannot be used to identify parents, but it can be used to identify the spouse and child of the LSM. The son and daughter-in-law of the head of household may be the LSM's parents, but they may be uncles or aunts - they would be allocated a different family number to the LSMs family at census processing, so it is not possible to tell which they are.

Because it may sometimes be impossible to identify a particular relationship, users may have to restrict their analyses to those cases where the LSM is in the head of household's family. However, one of the limitations of this is that it excludes those families which are more likely to be concealed - for example, young lone parent families.

In 1991, there is one additional situation where excluding concealed families still means that the use of the LSRELAT variable to identify parents will not be immediately obvious. Where a child is in a family where the parents are co-habiting, the following possibilities exist for coding the relationship between the parents and child:

Example 4: Children of cohabiting parents

1991 Census Schedule: (Relationship to head)	RELAT	LSRELAT	FANUM
Person no 1: Head	00=Head	07=Parent	1
Person no 2: Co-habitee	02=Living together as couple	26=Parent, other unrelated	1
Person no 3: Son	03=Son or daughter	LSM	1

Example 5: Children of a co-habitee

1991 Census Schedule:	RELAT	LSRELAT	FANUM
Person no 1: Head	00=Head	16=Other unrelated	1
Person no 2: Co-habitee	02=Living together as couple	07=Parent	1
Person no 3: Unrelated	04=Child of cohabitee of hoh	LSM	1

For both these examples, the census coding identified the family as one where the parents are co-habiting. The different coding for "parents" in each case reflects the relationship of the child to the head of household. In example 5, s/he is most likely a step-child of the head of household, while in example 4, the child is definitely a child of the head of household. Where it is unclear whether the LSM is a child or step-child of the cohabitee of the head of household, the family number (FANUM) can be used, as the co-habitee family will have been identified as such by the census coding. The 1991 Census was the first census in which co-habiting couples would have been allocated the same family number at census processing.

A similar situation occurs for the 1981 Census-LS data for those families where there is a *de facto* spouse (see *LS User Manual* p. 294). Where the relationship to the head of household was recorded as *de facto* spouse and a child was identified as the son or daughter of the head of household, the relationship between the child and the *de facto* spouse is coded as unrelated - when in fact this may not be case. Alternatively, if the child was unrelated to the head of household but was the child of the *de facto* spouse, s/he would be coded as unrelated to the head of household but the relationship between the parent and child would not be identified.

1981 *de facto* spouse families

Example 6: Child of head of household

Census Schedule:	HRC	LS RELAT	FANUM
Person no 1: Head	01=Head	04=Parent	1
Person no 2: de facto spouse	14=de facto spouse	17=Other unrelated	6*
Person no 3: Son	03=Son or daughter	LSM	1

*not in a family

Example 7: Child of de facto spouse

Census Schedule: (Relationship to head)	HRC	LS RELAT	FANUM
Person no 1: Head	00=Head	17=Other unrelated	6*
Person no 2: de facto spouse	14=de facto spouse	17=Other unrelated	1
Person no 3: Unrelated	16=Unrelated	LSM	1

*not in a family

In example 6, it would be sufficient simply to identify the *de facto* spouse of the head of household as a "parent". By comparison, in example 7 it would be necessary to find the head of the LSM's family, and where they are a *de facto* spouse of the head of household take both the head of family and head of household as parent figures. In both cases, the sex variable from the LS Core File would be used to identify the mother and father. This would enable inter-censal comparisons to be made between parents in order to identify both parent figures where present and whether or not they were the same as at the previous census points.

It should also be noted, that in a very few cases although families are identified at census there may be sex discrepancies which makes it impossible to identify the father or mother of an LSM.

Family type and relationships

The census definition of a family includes grandparent(s) and children where there are no apparent parent(s) present. Such cases can be identified where the head of household is the head of family and the relationship to head of household is grandchild.

At the 1971 and 1981 Census the household composition algorithm failed to identify lone parent families where the child(ren) was the grandchild of the head of household and the parent was the never married son or daughter of the household head. In these cases the three generations were assigned the same family number. A study of a sample of census forms from the 1971 Census carried out by OPCS concluded that

90% of these households contained a "concealed lone parent family". An additional problem in these households is that where the LS member is in the second or third generation it is impossible from the available information at census to code parent/child relationships to the LSM correctly. A person coded as parent may in fact be an uncle or aunt and a son or daughter may be a nephew or niece. Should these cases be relevant to your research it is possible to identify them when creating an extract from the Model 204 database. For a more detailed discussion of household composition data from the 1971 and 1981 Census see *LS User Guide No 1, Households, families and fertility (*Penhale, 1990).

To conclude, the LS may be used to identify specific relationships and circumstances within the household. Care should be taken to request all the necessary variables to identify the relationships required, and consideration should be given to the extent to which searching for the non-LS members is necessary for a particular research project. While LS variables are derived from a set of variables taken directly from the census schedule, further derivations may be necessary to maximise the benefit from both.

5 The LS and Travel-to-Work Areas (TTWAs), Mike Coombes, CURDS [3] Newcastle University

A brief item in the previous issue of this newsletter (Dodgeon, 1995) noted that there have been some delays in the availability of 1991-based Travel-to-Work Areas (TTWAs). This paper provides some background and then sets out the current position. It also provides an overview of the process of defining TTWAs. Finally, a number of alternative ways of linking TTWAs to LS data are described.

Why are TTWAs of interest?

One of the crucial attractions of the LS is that data are held at the individual level, rather than being pre-aggregated to areas like the local authority districts used for almost all other census outputs. Even so, many researchers require LS members to be coded by area, for example, because they are interested in the possible influence on people's well-being of the "context" in which they live, either at the neighbourhood or the locality level. The former may relate to "lifestyle" issues such as smoking or other health-related behaviour norms in higher or lower status neighbourhoods. (The LS includes an ACORN-type variable which may be used for this type of analysis.) The latter typically groups together all the contrasting neighbourhoods within a single town or city and its surrounding rural areas.

The research interest in localities often centres on labour market issues (eg the level of unemployment in a given area), although this can also spill over into more social and environmental concerns such as the impact of the area's local industries on pollution levels or the workforce's class structure. For example, an area once dominated by textile manufacturing is likely to have a legacy of certain long-term illnesses, a long tradition of high levels of female labour force participation with low wage levels.

There are many LS researchers who are keen to use a suitable form of locality coding to investigate these types of issue. This will need to be based on the labour market theme at the heart of the locality concept, but ideally would also reflect other aspects such as the social and environmental dimensions mentioned above (Coombes et al, 1988). It is obvious that local authority areas are far too arbitrary and inconsistent in their definitions to meet this need. For example, it would not be plausible to argue that the "context" for people living in Worcester is limited to the city's narrowly defined boundary. The effect of using local authority area in this way is that the "context" for people living just outside the city in the adjacent suburbs and villages does not include any reference to the city's jobs, services or other accessible features. Another factor to take into consideration is the inconsistent way in which local authority areas are defined. For example, the boundary of Carlisle, a similar-sized city to Worcester, extends well beyond its built-up area to include a very wide area of northern Cumbria. Thus district boundaries are highly inconsistent, and so a proportion of them will be inappropriate for any study which needs to represent the locality "context" in a meaningful way.

How are TTWAs defined?

TTWA boundaries provide a set of areas within which most commuting flows are self-contained (ie relatively few commuters live in one TTWA but work in a different one). Since 1983, TTWAs have become unique among official boundaries in that they are derived directly from research undertaken by academics (Dept of Environment, 1984). One reason for this is that the TTWAs' value depends on them being based on statistical and scientific criteria, and the fundamental academic inputs provide rigorous independent "quality control" to the definitions.

---

[3] Curds is the Centre for Urban & Regional Development Studies: the TTWA research was undertaken by the North East Regional Research Laboratory (NE.RRL) team, led by Mike Coombes in conjunction with Colin Wymer (Planning Dept.) and Stan Openshaw (Leeds University).

The primary use of TTWAs is in the reporting of unemployment rates. One initial objective was to have as many TTWAs as possible, to maximise the detail of information which is published (Dept of Employment, 1984). Complex statistical factors, relating to the way in which the unemployment rate data are compiled, impose other criteria:

* TTWAs should not be so small as to make them statistically very volatile (although the minimum population was below 10,000); **and**

* TTWAs must satisfy a self-containment minimum (previously this was approximately 70%).

The technical challenge arises from the fact that the analysis has to process a matrix of commuting flows between over 11,000 wards across the whole of Britain. The iteration process used to produce a set of TTWA boundaries requires very sophisticated data storage, retrieval and manipulation software. The basic algorithm and software development for the TTWA definitions was devised by the Newcastle team when the current set of TTWAs was created in 1983/4 (Openshaw et al, 1988).

Why are TTWAs revised each decade?

The TTWA definitions have been updated over the last two decades to meet the government's need for accurate and consistent definitions of labour market areas. The factors which contributed to the increase in average commuting distances between 1981 and 1991 include:

* the loss of traditional industrial jobs, which were often staffed locally,
* the continuing dispersal of population,
* the increasing affluence of the population, **and**
* further growth in car usage.

Of course, there were also some countervailing pressures, such as the strong growth in part-time working and further increases in the female workforce - both categories of workers who tend to have shorter commuting distances. The suburbanisation of some types of job could have reduced commuting distances too, although experience in the USA suggests that this is unlikely. Figure 3 shows the commuting flows of more than ten workers for England and Wales. (This figure is not available in the .PDF version of *Update*.)

The key measure in the definition of a Local Labour Market Area (LLMA) is the self-containment of commuting. As commuting distances between 1981-91 have lengthened, more commuters tend to be crossing the LLMA boundaries. A first broad estimate is that self-containment values for TTWAs have declined by an average of about five percentage points. This is a slightly faster decrease than during the period 1971-81. For example, a fairly balanced LLMA in 1971 might have been 75% self-contained - that is, only a quarter of its workforce commuted across the boundary. This value would have declined to just over 70% in 1981 and is approaching 65% by 1991. In other words, in 1991 over a third of the workforce were commuting across the boundary.

Clearly, these broad generalisations disguise considerable variation between areas. Similarly, while all sub-groups of the workforce seem to be travelling further to work, this change may be much more marked for some. For example, less than half (44%) of men in 1991 commuted no more than 5 kms, whereas the values for women were 53% for those with full-time jobs and 70% for women in part-time work. This substantial contrast in 1991 follows a period when underlying gender differences had fallen. In particular, the gap in car usage for commuting has narrowed from 1981 when 59% of men but only 37% of women used cars, to 1991 when the equivalent values were 67% of men and 52% of women.

Notwithstanding these 1981-91 trends, the contrasts between different sub-groups in the self-containment of commuting flows is still probably larger than is often realised. Therefore when any new set of TTWAs is defined from an analysis of all workers' commuting flows, it should be remembered that a very complex pattern of different group-specific sets of boundaries has been simplified in order to provide a definitive single set of areas.

To summarise, TTWAs are the official set of British local labour market area definitions. They are of interest for studying those local contextual effects which extend beyond a single neighbourhood. However, TTWAs are not the only possible set of labour market area definitions - as shown, for example, by Green et al (1986). The danger is that TTWAs are often used "because they are there" rather than because they are ideal for a particular line of research into contextual influences.

TTWAs and 1991 data

While TTWAs have been used in the past simply "because they are there", ironically they are not yet "there" for the 1991 Census data. Here it is important to separate out two issues:

* the definition of new 1991-based TTWAs **and**
* the coding of 1991 data in the LS to current TTWAs (1981-based).

The position on the first issue is still unclear. The NE.RRL team carried out the research to produce the 1991-based TTWA boundaries for the Employment Department and delivered these on schedule in Spring 1994. After the Employment Department was abolished in July 1995, responsibility for labour market statistics was assumed by the Central Statistical Office (CSO). A recent statement from the CSO notes that "the issue of reviewing the TTWAs on 1991 data is currently being revisited from a CSO perspective" (Everett, 1995). In the short term, then, there are no available up-to-date TTWA definitions.

During this interim period, the NE.RRL has been commissioned by the ESRC to produce definitions of localities designed to be of greater interest to the academic community. These new areas will reflect not only commuting patterns but also other localised movements, such as migration and the use of some services such as banks. These locality definitions will soon be completed and should be accessible early in 1996 from the MIDAS national computing facility at Manchester University.

The second issue begs another question, which may have a different answer for each LS researcher thinking of using TTWAs. This question is whether a new set of boundaries are needed or whether another set of definitions would suffice ? For example, the 1981 TTWAs include Rotherham and Sheffield as separate areas; but it is quite possible that the trends in local employment levels and commuting patterns will lead to the two being combined. (This is not a specific prediction, merely an illustrative example.). Rotherham is a locality which has been dominated by coal mining, whereas Sheffield was traditionally a steel making city which is now a major service centre. If they are combined then the two areas' unemployment rates, and other data used to represent the "context" of the LS members, will be assumed to be the same in both localities. (In this event Sheffield's characteristics would dominate, due to its larger population). For studies of longer-term processes, such as the effect of local industrial structure on health or migration behaviour, it seems that it would be better to keep the two localities separate. More significantly, for a longitudinal analysis which uses 1981 (or earlier) LS data, it would be more appropriate to use the TTWAs which reflect the relative distinctions of the two localities at that time. In other words, the 1981-based TTWAs may well often be the better definition of "context" for longer-term analyses.

How can the 1991 LS data be coded to 1981-based TTWAs so that the locations where people are now living can be aggregated to the appropriate set of locality boundaries for longer-term studies? This is a purely technical problem, arising from the changes in the "building block" areas used for coding 1981 and 1991 data. Fortunately, there are ways around this problem. A "look-up" file coding 1991 areas into current (1981-based) TTWAs will soon be available on the NOMIS database at Durham University (Everett, 1995). Another route, is to use a "look-up" file, created by the NE.RRL, already available via MIDAS at Manchester University (in the "Useful Files" facility). The second file provides a link between the 1981 and 1991 "building block" areas, from which the 1981 definition of TTWAs can be used to identify the relevant current (1981-based) TTWAs within which each 1991 Census ED (enumeration district) is located. Each LS member is already coded to the ED in which they were living in 1991, so this series of linked files provides the basis for also defining their "context" in terms of existing TTWA boundaries.

So LS researchers may soon have a number of options for characterising each LS member's locality. For example, it may be possible to represent the context in which an individual lives in terms of new (1991-based) TTWAs, or of the new ESRC-sponsored locality definitions. For longitudinal work, the use of 1981-based TTWAs may be preferable. If such a range of options can be provided, the onus will be on each individual researcher to decide which is the most appropriate form of locality definition - rather than deciding whether to use TTWAs simply because "they are there" or, perhaps, because they are not!

References:

Coombes M G ,Green A E and Owen D W (1988) "Substantive issues in the definition of localities: evidence from sub-group local labour market areas in the W. Midlands" , *Regional Studies,* **22**, pp 303-318

Dept of Employment (1984) "Revised travel-to-work areas", *Employment Gazette Occasional Supplement 3*

Dodgeon B (1995) "1991 travel-to-work areas", *Update- News from the LS User Group,* **Issue no. 11**, page 3

Everett G (1995) "New travel-to-work areas", *NOMIS News,* **10**, p 5

Green A E ,Coombes M G and Owen D W (1986) "Gender-specific local labour market areas in England & Wales", *Geoforum,* **17**, pp 339-351

Openshaw S, Wymer C and Coombes M G (1988) "Making sense of large flow datasets for marketing and other purposes", *North East Regional Research Report 88/3,* CURDS, University of Newcastle upon Tyne.

6 Opportunities and pitfalls of the 1991 Census labour market data

An article by Catherine Hakim published in the September 1995 issue of *Work, Employment and Society* alerts LS as well as SAR users to a feature of the labour market data in the 1991 Census that is of great importance for the design of research analyses and for the interpretation of results. Changes in census question-wording in recent decades have led to retrospective labour market data being collected as well as data on the current workforce. Inclusion of the retrospective data significantly alters the composition and characteristics of the workforce, as reported by the 1991 Census. The retrospective information also provides an extra dimension to census data, which can be exploited for analyses of labour mobility. While the article focuses on the two Samples of Anonymised Records (SARs) now available via the Cathie Marsh Centre for Census and Survey Research (CSSR) at Manchester University, the issues which it addresses apply equally to analyses based on the 1991 Census data now available in the LS.

The paper describes the small incremental changes in question-wording in the 1971, 1981 and 1991 Censuses that led to labour market data being available for two substantially different bases in the 1991 Census microdata, allowing users to choose a wide or narrow concept of the labour market. Analyses of the 1% Household SAR and the 2% Individual SAR are presented to show how the composition and characteristics of the workforce will differ, depending on how widely or narrowly the boundaries of the labour market are drawn. For example, female-dominated occupations may be artificially expanded or contracted. The article highlights that researchers analysing 1991 Census microdata will have to choose (and justify) their choice of 1991 base population carefully. The paper uses illustrative examples, to show how this new feature of the census data can be used as a rough approximation to labour mobility data for the inter-censal decade, opening up a wide range of analyses not previously possible. In the case of the LS, which allows comparisons between occupation and labour market status in 1981 and 1991, the additional information for the inter-censal period makes for a richer dataset. For example, information on an

individual's last occupation in the inter-censal decade can be used in analyses of their current characteristics and situation, such as divorce and remarriage rates (Hakim, 1994, page 446).
Despite the design differences between the two SAR samples, they seem to yield identical results on the labour market. This suggests that the 1991 Census 1% sample data in the LS would also yield a similar workforce profile, allowing the LS and the SARs to be used inter-changeably and in combination, according to the nature of the labour market and other variables required for a study.

References:

Hakim C (1994) "A century of change in occupational segregation 1891-1991", *Journal of Historical Sociology*, **Vol 7 (December 1994)**, pp 435-454

Hakim C (1995) "1991 Census SARs: opportunities and pitfalls in the labour market data", *Work, Employment and Society*, **Vol 9:3 (September 1995)**, pp 569-582

We will be delighted to receive your comments on any of the articles in this newsletter.