CENSUS ANALYSIS: PLANNING THE WAY AHEAD

A Report on the Launch Seminar of the Census Analysis Group
April 6/7th 1992

Edited by Philip Rees

WORKING PAPER 92/13

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

Working Paper 92/13

CENSUS ANALYSIS: PLANNING THE WAY AHEAD

A Report on the Launch Seminar
of the Census Analysis Group

April 6/7th 1992, University of Leeds

Edited by

Philip Rees
School of Geography
University of Leeds
Leeds LS2 9JT

with contributions from:

Philip Rees, University of Leeds
Ian Mills, Census Division, OPCS
Keith Cole, University of Manchester
Tony Champion, University of Newcastle-upon-Tyne
Daniel Dorling, University of Newcastle-upon-Tyne
Mike Coombes, University of Newcastle-upon-Tyne
Bob Barr, University of Manchester
Alex Hirschfield, University of Liverpool
Peter Brown, University of Liverpool
John Marsden, University of Liverpool
David Owen, University of Warwick
John Giggs, University of Nottingham
Cathie Marsh, University of Manchester
Elizabeth Middleton, University of Manchester
Robin Flowerdew, University of Lancaster
Paul Boyle, University College of Swansea
John Stillwell, University of Leeds
Bob Woods, University of Liverpool
Paul White, University of Sheffield

The Census Analysis Group (CAG) is an informal group of researchers at Higher Education Institutions with an interest in investigating the social and spatial structure of Britain's population and society using data from the 1991 Census of Population. The CAG is sponsored by a seminar grant from the Census Initiative of the Economic and Social Research Council, the Information Systems Committee of the Universities Funding Council and the Department of Education, Northern Ireland. The life of the CAG is currently scheduled for calendar years 1992 and 1993.

CONTENTS

1. Introduction

Philip Rees, University of Leeds

A two day meeting was held at the University of Leeds to discuss the ways in which researchers at higher education institutions were planning to use forthcoming Census data in their research. The seminar was designed (1) to provide information on the Census output products and the timing of their dissemination, (2) to provide a forum for researchers to brainstorm about their future analysis plans and (3) to stimulate the creation of networks of collaborating scholars interested in particular features of the Census results.

Presentations at the seminar were mainly oral. This document provides a record of the meeting in the form of a post-conference report. The report is an edited version of papers and material presented, for which the editor takes responsibility. Further details of each individual presentation can be obtained from the speakers.

It is hoped that new plans for analysis will be currently emerging that can be presented as papers at later meetings of the Census Analysis Group. The next meeting of the Census Analysis Group will be held at the Annual Conference of the Institute of British Geographers on Wednesday January 6th 1993 at Royal Holloway and Bedford New College at Egham, Surrey. The theme is to be *Small area population change in the UK: 1981-91*. The session is co-sponsored by the Population Geography Study Group.

2. The 1991 Census, the Census Analysis Group and the ESRC/UFC-ISC 1991 Census of Population Initiative

Philip Rees, University of Leeds

Full details of this presentation are given in a paper entitled "The ESRC/UFC-ISC 1991 Census of Population Initiative: Delivering the Data of the Decade" to be published in the ESRC Data Archive Bulletin. Copies of drafts of the paper are available from the author. Here a summary is provided.

In order to carry out analysis of the results of the 1991 decennial Census, it is necessary to have access to census datasets in as convenient a form as possible. This is the goal of the Census Initiative funded and sponsored by the Economic and Social Research Council (ESRC), the Information Systems Committee of the Universities Funding Council (UFC-ISC) and the Department of Education, Northern Ireland (DENI). ESRC and UFC-ISC are contributing £1.5 millions each and DENI some £0.12 million, while ESRC are responsible for all data, training and research contracts.

The Census Initiative consists of some seven components: (1) the Census datasets, (2) the digitised boundary and locational reference datasets, (3) the units for disseminating the datasets, (4) the training programme, (5) the development programme, (6) the seminar programme and (7) the coordinator. Each of these are considered in turn.

2.1 *The Census datasets*

There are some five datasets which will be provided in machine readable form.

2.1.1 The Small Area Statistics (SAS) and Local Base Statistics (LBS)

These are currently (June 1992 to November 1992) being provided for English and Welsh counties and for Scottish regions. The SAS will be available for enumeration districts (EDs) in England and Wales and for output areas in Scotland (about 130,000 units of about 200 households each). The LBS will be available for wards in England and Wales and postcode sectors in Scotland. OPCS's User Guide 3 gives details of the layout of the 100 or so tables of statistics which are provided in full (LBS) or abbreviated (SAS) form.

The SAS and LBS will be provided for all standard higher geographies when these become available, so as to avoid the repetition of aggregation operations. The SAS and LBS for England and Wales will be provided for wards, districts, counties, standard statistical regions, the nation, Parliamentary Constituencies, European constituencies, District Health Authorities and Regional Health Authorities. The SAS will also be available for postcode sectors and civil parishes/communities in Wales. The

SAS and LBS for Scotland will be made available for postcode sectors, Districts, Scottish regions, Parliamentary constituencies, European constituencies and Health Board Areas. The SAS will be provided for civil parishes.

The Northern Ireland Small Area Statistics will not become available until 1994 and will be a much smaller set of tables than those for Great Britain. They will be available for grid squares (100 metre squares in urban areas, 1 kilometre squares elsewhere).

2.1.2 The Samples of Anonymised Records (SARs)

These are a 2% sample of individuals and a 1% sample of households drawn from the 1991 Census enumeration returns for analysis by the academic community. In particular, researchers will be able to design their own crosstabulations. Fuller details are given in section 12 of this report. The datasets will be held at the University of Manchester and made available by the Census Microdata Unit.

2.1.3 The Longitudinal Study (LS)

This is a sample of individuals who have one of four selected birthdays taken from the Census of 1971 of England and Wales linked through to the 1981 Census, and maintained through addition of births and immigrations and through subtraction of deaths and emigrations, for England and Wales. During 1993 the characteristics of LS members traced in the 1991 Census will be added. The data reside in datastreams held on the main OPCS computer and are accessible to academic users via the LS Support Programme at the Social Statistics Research Unit (SSRU), City University. Full details of the LS and access procedures are given in SSRU (1990).

2.1.4 The Special Migration Statistics (SMS)

These statistics provide a limited number of counts for migrants and migrant households over the year prior to the 1991 Census on the 21 April 1991).

There will be three SMS datasets supplied by the Census Offices to the CDU at Manchester Computing Centre for online access. SMS1 will provide flows between wards (England and Wales) or postcode sectors (Scotland), giving the sex and broad age group of migrants, and the total number of wholly moving households and residents in those households. SMS2 will provide a table of inter-district migration classified into 10 simple tables (plus an additional table by language in Wales and in Scotland). SMS3 will contain summaries of the inter-district tables by type of in- and out-migration. Both SMS2 and SMS3 will be subject to confidentiality constraints, although Tables 1 to 3 giving the demographic structure of migrants will always be available.

The Census Initiative is negotiating the supply of an im-

proved 1991 version of MATPAC, the program for data extraction.

2.1.5 The Special Workplace Statistics (SWS)

This dataset will contain journey to work statistics on a ward/postal sectors geography and will be held for general academic access at MCC via MATPAC91.

2.2 *The locational reference datasets*

Users need to locate the census data they use for the purposes of reference, mapping or integration with non-Census datasets. A variety of locational datasets are to be provided.

The *Census codes* for all small areas are embedded in the SAS/LBS datasets in the form of the Ordnance Survey coordinates of centroids assigned by the Census Offices.

Indexes and directories of census and postal codes are being procured. The Census Initiative is purchasing for Scotland an *index file* relating Output Areas and higher geographies to postal codes. The equivalent products for England and Wales are the *Area Constitution File* and the *ED/Postcode Directory*, which are being purchased.

The Census Initiative is ordering ten microfilm copies of *ED Planning Maps* on a 1:10,000 OS Grid background for England and Wales, and investigating the provision of transparent overlays of Output Area boundaries suitable for use with OS maps for Scotland.

The Census Initiative is purchasing *digitised boundary datasets* that provide a numerical representation of the line segments that, when stitched together by suitable software, describe the boundaries of the smallest geographical building blocks used in the 1991 Census. For Scotland a full set of segment descriptions of the unit postcodes which were used to build Output Areas are being procured in the proprietary GIMMS format, from which boundary sets for the higher geographies can be formed. For England and Wales, the Census Initiative is purchasing a full set of ED boundaries in generic segment format from the EDLINE Consortium. These ED datasets are being subjected to Quality Assurance checking at the University of Manchester. The ED boundaries will be dissolved to form higher geography datasets and utilities will be provided to convert the data from generic format into the common proprietary formats for mapping.

As a result of these planned purchases and developments, researchers should be able to reference, manipulate and map 1991 Census SAS/LBS data with much greater facility than at previous censuses.

2.3 *The units*

The purchase of datasets themselves does not guarantee that

academic researchers will be able to use the data easily and effectively. The datasets need installation on suitable computers, detailed easy-to-use documentation should be provided and a help service for users must be available. The user support needed will be provided by two existing ESRC funded units and two new units funded by the Census Initiative.

The ESRC Data Archive at the University of Essex will act as the depository for all Census datasets and will provide descriptions of each Census dataset on their powerful Bibliographic Information Retrieval ONline (BIRON) system. Researchers wishing to link Census data with data from large national surveys will find BIRON particularly helpful prior to negotiating access to suitable datasets.

The LS Support Project at the Social Science Research Unit, City University, will provide access to the updated LS that will include a sample of linked records from the 1971, 1981 and 1991 Censuses together with intermediate vital events. Full details of the LS are given in SSRU (1990).

The Census Microdata Unit at the University of Manchester is being funded by the Census Initiative to house, disseminate and act as a research focus for the SARs datasets. A smaller project to be located at the University of Leeds will develop versions of the SAR datasets suitable for use on UNIX workstations.

The Census Dissemination Unit at the Manchester Computing Centre will play a major role in providing access for the academic community to the SAS/LBS, SMS, SWS and digitised boundary datasets. The Census Offices will supply the data on magnetic tape to the CDU for installation on MCC's CMS service where users will be able to use the access program, SASPAC91. Subsequently, copies of the datasets will be shipped to other Higher Education sites which wish to provide a local service.

2.5 *The training programme*

The Census Initiative is funding six projects that will provide training for users of the 1991 Census over and above the materials and documentation provided by the Census Offices and the Units.

A Guide to the 1991 Census is being edited by two unit directors (Angela Dale and Cathie Marsh). The Guide, to be published by HMSO in early 1993, will provide the essential information for users to plan and to guide their use of 1991 Census data. Chapters will cover: fieldwork and data entry, confidentiality concerns, Census coverage and quality, published census data, the SARs, the Longitudinal Study, the Small Area Statistics, migration and transport statistics from the Census and the follow-up surveys.

Tutorial and user guides for the 1991 Census of Population data, being prepared by a team (Martin Clarke, David Martin and

Neil Wrigley) at the GeoData Institute (University of Southampton), will provide a self learning package consisting of a set of documents and a PC diskette containing the tutorial notes and illustrative data, applications and programs, including guides on using the ED/Postcode Directory and matching 1991 Census SAS with other datasets.

Sample datasets to illustrate different applications are being assembled by a Midlands Regional Research Laboratory (MRRL) team (Allan Strachan, David Walker and H.M. Hearnhsaw). Topics to be covered range from ethnicity at small area scale linked to OPCS fertility statistics to economic activity and employment statistics at standard region scale linked to Labour Force Survey data.

A second team (Mitchell Langford, David Unwin and Allan Strachan) at the MRRL will develop *A trainer's resource base for using the 1991 Census*. This will consist of a training package (lecture course materials) for instructors in Higher Education who wish to give courses on Census data, modelled on the package developed for training users of the UNIRAS computer graphics package.

A faculty team (Ian Bracken, Teresa Rees and Stephen Littler) at the University of Wales College of Cardiff is organising *a series of training workshops* targeted at census users in planning, policy analysis and in local government covering the topic areas of housing, health, employment and localities.

The final component of the training programme will be a week long intensive workshop planned by the Census Microdata Unit, the Census Dissemination Unit and the LS Support Programme (Cathie Marsh and Angela Dale) to be held in the Summer of 1993 at the University of Manchester. The course is aimed at teachers of research methods courses in British Higher Education Institutions. A key feature will be hands-on access to the Census datasets described. The course will be repeated if demand warrants.

2.6 *The development programme*

The Census Initiative seeks to encourage the development of new uses of Census data, the production of new added-value products based on census data and the investigation of common methodological problems. Some six development projects are due to start by the beginning of 1993.

A project on *Interpretative analysis of social change using spatial models of Census data* with principal investigators in Southampton and Cardiff (David Martin and Ian Bracken) aims to apply surface modelling techniques to overcome problems of comparability in using areas of different sizes and shapes to map and interpret census data.

A team (Tim Holt, Neil Wrigley) from Southampton will inves-

tigate *A new approach to the "aggregation" issue in Census data analysis* so that a practical technique for adjusting the outcomes of analyses based upon spatially aggregated areal data to yield estimates of the underlying individual level parameters.

A research project (directed by Peter Elias) at the University of Warwick aims to provide information on *Occupational definition and coding conventions for the 1981 and 1991 Censuses*, which will provide the researcher with the means of distinguishing classification change from real change in occupational distributions.

An investigation is planned by a University of Manchester researcher (Bob Barr) of *American Census data handling techniques and the British Census* with a view to assessing the applicability of the techniques and software used in the USA to redistrict electoral areas using Census data.

A team (Philip Rees, John Stillwell, Mark Birkin and Debbie Phillips) at the University of Leeds plans an *Analysis of migration flows* measured by the 1991 Census and a *Simulation of whole populations with added income attributes*. The first sub-project will make a systematic comparison of Census and NHS Register migration data in order to combine the datasets in local population estimation and projection, while the second project aims to link Census and large national survey datasets using microsimulation in order to add income attributes to census small area datasets.

Finally, a research team (Stan Openshaw) relocating to Leeds aims to carry out a detailed empirical study of the magnitude of scale and zoning effects on the inference of association between individual attributes from spatially aggregated data. Best practice advice on how to ameliorate or avoid *aggregation problems* would be offered.

The reader will probably note that this project, the previous one and the Southampton project intend to exploit the availability of both aggregate data (SAS/LBS) and individual data (SARs) for the first time.

2.7 *The seminar programme*

The Census Initiative is funding a seminar programme, which is organised by the *Census Analysis Group* with a steering committee consisting of Tony Champion, Philip Rees, John Stillwell, Paul White and Bob Woods. This is a limited life network of researchers, including all initiative funded researchers. The seminar programme aims to provide a forum for the discussion of the outcome of analyses of the results of the 1991 Census. The first meeting was held in April at the University of Leeds. A half day session is planned at the next Annual IBG meetings (January 1993) at Royal Holloway and Bedford New College and a three day conference is in preparation for September 1993 at the University of Newcastle.

2.8 *The coordinator*

Given the size and complexity of the programme of dataset acquisition, support, training and development outlined above, the Census Initiative needs co-ordination. This is a job to which the Economic and Social Research Council has seen fit to appoint Philip Rees. In essence, the coordinator's task is to ensure that the national research resource which is provided by the Census Offices reaches all those researchers who would benefit from using census type data to paint a fuller socio-geographic portrait of the British population at the start of the century's last decade and to stimulate publication and dissemination of the analyses of the results of the 1991 Census of Population.

3. Current output programme of the Census Offices

Ian Mills, Census Division, OPCS

The presentation concentrated on two topics: (1) the reasons for the delay in producing Census results and (2) the likely timetable for those results.

The Census Office discovered a serious problem in the Census data which has forced rescheduling of the entire output programme. This problem, if uncorrected would have led to many people being classified as students rather than as employed, self-employed or unemployed. This misclassification, which applied to nearly half a million people, would have seriously impaired the usefulness of the Census statistics. Work has been underway since December 1991 to deal with the problem.

The problem involved the way in which question 8 on the Census Questionnaire Schedule was filled in. About 500,000 people ticked the box "This address" when asked for "Term time address of students and schoolchildren" when other information on the form clearly indicated they were not a student or schoolchild. The Edit system was incorrectly designed to deal with this problem and had to be revised. To correct the error it was necessary to go back to the raw data and extract the correct economic activity details. A test has been applied to determine whether a particular record belonged to a student or not, and then for non-students the original response to the Question 13 on economic activity in the week before the census was reinstated and the answer to the term-time address question amended.

The *timetable* for Census output has had to be substantially revised Priority has to be given to the provision of statistics for the re-basing of the mid-year population estimates of the Registrar Generals and for resource allocation work in the Department of Health and the Department of the Environment. The revised output timetable is displayed in Table 1 (this is an updated version of that presented at the Seminar in which one month has been added to each item).

There will be some *residual statistical effects* of the question and processing problem. Counts of employees, the self-employed, the unemployed and those on Government schemes will be as recorded on the Census forms. Counts of economically active students will be slightly too low. Counts of economically inactive students will be very slightly low. Counts of those retired, permanently sick, looking after the home or family, or other economically inactive will be very slightly high. Students with term-time address "elsewhere" will be as recorded on the Census forms while students with term-time address "at this address" will be slightly too low. These effects will be quantified in the reports of the *Census Validation Survey*, Part 1 of the report on which will appear in Summer 1992 and Part 2 during 1993. It will be possible to estimate the degree of underenumeration in the Census from the Census Validation Survey. The Preliminary Report statistics, based on enumerators' returns,

were estimated to be about 2% underenumerated. As a result of late returns of missing forms and imputation procedures for non-returns from occupied households, it was anticipated that underenumeration would be less than 1% for the statistics to be published during 1993 in a series of *National Reports* on the topics outlined in Table 2.

Table 1. Revised output timetable

County/Region Monitors	
Publication of first counties/regions	May 1992
final counties/regions	Oct 1992
100 per cent LBS/SAS for primary areas	
Supply of first counties/regions	June 1992
final counties/regions	Nov 1992
County/Region Reports (Part 1	
Publication of first counties/regions	July 1992
final counties/regions	Jan 1993
National Summary Reports (Part 1)	
Publication of reports for GB, Scotland and Wales	Mar 1993
ED/Postcode Directory (England and Wales only)	
Supply of first counties	June 1992
final counties	Nov 1992

Table 2. Topic reports planned for the 1991 Census

No.	Title	Processing base
1.	Historical Tables	100%
2.	Sex, Age and Marital Status	100%
3.	Usual Residence	100%
4.	Persons aged 60 and over	100%
5.	Housing (including cars)	100%
6.	Communal Establishments	100%
7.	Household Composition	100%
8.	Limiting Long-Term Illness	100%
9.	Welsh Language in Wales	100%/10%
10.	Gaelic Language in Scotland	100%/10%
11.	Ethnic Group and Country of Birth	100%/10%
12.	National Migration - Part 1	100%
13.	National Migration - Part 2	10%
14.	Regional Migration - Part 1	100%
15.	Regional Migration - Part 2	10%
16.	Economic Activity	10%
17.	Workplace and Transport to Work	10%
18.	Household and Family Composition	10%
19.	Qualified Manpower	10%
20.	Children and Young Adults	100%/10%

4. 1991 Census support and dissemination arrangements: MCC's role in the ESRC/UFC-ISC initiative

Keith Cole, Manchester Computing Centre

Manchester Computing Service (MCC) are a provider of national services to the academic community in the form of user support, networking and software distribution services. MCC provides access to major national databases, which include the 1981 Census of Population SAS, SMS and SWS datasets, OPCS Vital Statistics for Wards (1981-89), the Post Office Central Postcode Directory, the 1981 UK digitised Boundary Data Base (wards and postcode sectors), General Household Survey, Family Expenditure Survey, National Labour Force Survey and National Child Development Study.

To these datasets will be added the 1991 Census datasets for England and Wales (Table 3), Scotland (Table 4) and Northern Ireland (details being negotiated). The Tables show the areal levels for which SAS and LBS data are to be available. The LBS contain some 99 tables or circa 20,000 counts and are provided down to ward/postcode sector level. The SAS constitute and abbreviated version of the LBS with 86 tables and circa 10,000 counts, available down to enumeration district/output area level. For Northern Ireland the SAS will contain circa 850 counts and be provided for grid squares.

The SAS and LBS are sets of crosstabulations of the populations of individuals and households resident in geographic areas. The difference between the detail provided by the two sets can be illustrated by looking at the tables for ethnic group to be provided. The ethnic question was one of the four new questions asked at the 1991 Census, the others being concerned with long term limiting illness, term-time address of students and the availability of central heating. The LBS will contain (Table 6) a crosstabulation of ethnic group by five year age group for each sex, while the corresponding SAS table will provide separately the crosstabulation of ethnic group by sex and ethnic group by five broad age groups.

The ESRC/UFC-ISC Census Initiative is providing funding for five years of a 1991 Census Support Officer who together with other input from MCC staff will constitute the Census Dissemination Unit. The support and dissemination functions undertaken will include: (1) receipt of datasets from OPCS and installation on MCC's CMS service, (2) a national on-line data service, (3) data distribution to other academic community sites of whole sets or subsets, (4) documentation, (5) training, (6) support services, and (7) a service of information about the census.

Users will be interested in the sizes of the SAS and LBS datasets in order to assess the storage requirements for holding local copies. Please note that other academic community sites which wish to provide the data to their users must draw up an approved User Registration System, similar to that to be used at MCC. Table 5 contains estimates of likely file sizes.

Access to the SAS and LBS datasets is provided through the Small Area Statistics Package for the 1991 Census (SASPAC91) has been developed for local authority use by MVA Systematica and the London Research Centre. The Common Higher Education Software Team (CHEST) has arranged a site licensing system for use of the SASPAC91 software, which costs about £3,750 for five years for sites taking out licences now. So far, some 13 sites have taken out licences, but more are likely to do so as SAS and LBS become available. SASPAC91 is available for a variety of mini/mainframe operating systems, for PCs running DOS and for UNIX systems.

The Census Initiative is exploring the possibility of supplying SAS and LBS data on CD-ROM to the academic community, once the full datasets have been acquired and installed for use over JANET and some experience has been gained with its use.

The Census Dissemination Unit has set up an email distribution list for circulating news about census datasets. To join this circulation list, communicate with

CENSUS-REQUEST @ UK.AC.MANCHESTER

To circulate information to the list, send it to

CENSUS @ UK.AC.MANCHESTER

For more information, contact the author at

K.COLE @ UK.AC.MCC

or by telephone on 061 275 6066.

Table 3. ESRC/UFC-ISC 1991 Census of Population Initiative purchase of SAS, LBS and related materials for England and Wales

Areas	SAS	LBS
Enumeration Districts	+	
Wards	+	+
Districts	+	+
Counties	+	+
Standard Statistical Regions	+	+
National (England & Wales)	+	+
Parliamentary Constituencies	+	+
European Constituencies	+	+
District Health Authorities	+	+
Regional Health Authorities	+	+
Postcode Sectors	+	
Civil Parishes/Communities in Wales	+	
ED/PC Directories		
Area Constitution File		
Maps of EDs on microfilm - Ten sets		

Notes:
1. The + sign indicates that a dataset exists for that spatial level.

Table 4. ESRC/UFC-ISC 1991 Census of Population Initiative purchase of SAS, LBS and related materials for Scotland

Areas	SAS	LBS	SMS1	SMS2	SMS3	SWS
Output Areas	+					
Postcode Sectors	+	+	+			+
Districts	+	+		+	+	
Region	+	+				
Scotland	+	+				
Parliamentary Constituencies	+	+				
European Constituencies	+	+				
Health Board Areas	+	+				
Civil Parishes	+					
Postcode to Output Areas Index						
Output Areas to Higher Areas Index						
ED boundaries (digital)						

Notes:
1. The + sign indicates that a dataset exists for that spatial level.

Table 5. Local and Small Area Statistics: file sizes

Dataset	100% data			10% data			Total		
	ED/OA	WA/PS	DI	ED/OA	WA/PS	DI	ED/OA	WA/PS	DI
Storage requirements per area (Kilobytes)									
1991 SAS	31	35	58	20	23	38	51	58	96
1991 LBS	-	90	144	-	30	48	-	120	192
Example: file sizes for West Yorkshire (Megabytes) (4182 EDs, 126 Wards, 5 Districts)									
1991 SAS	126	4	.3	82	3	.2	208	7	.5
1991 LBS	-	11	.7	-	4	.6	-	15	1.3

Total storage requirement for SAS and LBS for Great Britain

Circa 6 Gigabytes

Notes:

1. ED = enumeration district
   OA = output area
   WA = ward
   PS = postcode sector
   DI = district

2. One bit = 1 binary digit
   One byte = 8 bits
   One kilobyte = 1000 bytes
   One megabyte = 1000 kilobytes
   One gigabyte = 1000 megabytes

5. The problems faced in measuring change and topic comparability between the 1981 and 1991 Censuses

Tony Champion, University of Newcastle-upon-Tyne

Most academic and social scientists will wish, when the full results of the 1991 Census become available, to explore the degree of change that has occurred at least since the last Census. There will be, however, a second set of questions that need to be addressed, namely, how do the trends between 1981 and 1991 compare with those between 1971 and 1981.

The following problems will be faced by any researcher trying to answer these questions with the aim of distinguishing artefactual change from real change. The list is derived in part from a Ph.D. dissertation written by Chris McGee of Birkbeck College, London in 1989.

(1) Change in Census coverage

It is likely that all Censuses are misenumerated to a small extent and that the level of misenumeration varies from Census to Census. The 1991 Census is likely on the basis of the Preliminary reports to have suffered slightly greater underenumeration that the Censuses of 1971 and 1981, principally because a small number of people wished to remain unknown to official statisticians because of a fear that census information would leak to Community Charge officials.

(2) Change in Census questions

It will not be possible to compare directly the answers to the new questions introduced in the 1991 Census, by definition. However, there will be a real desire to make backward estimates of ethnic group populations in 1981 and 1971, so that some idea of local change can be made. This topic is taken up later in the contribution of David Owen (section 10).

(3) Changes in variable definition

A careful study of the Census Definitions volumes from each Census will be essential. Among changes to look out for are the following.

(a) Different processing errors

(b) Differences in Census timing. The 1981 Census was held on the 6th April 1981 whereas the 1991 Census was held on the 21st April 1991, so that the intercensal interval is 10 years and 15 days not 10 years. Demographers modellers please note!

(c) The population base has changed in each successive Census.

(d) Change has occurred in the treatment of shipping counts.

(e) Changes have occurred in the enumeration of communal estab-

lishments.

(f) Changes have occurred in the definition of a room.

(g) New categories such as "cohabitant" (self-defined) have been introduced, compared with the earlier "de facto spouse" (imputed).

(4) Changes in Census geography

It is clear that comparable areas need to be defined in 1971, 1981 and 1991. In Scotland and in Northern Ireland the definition of comparable areas should be relatively straightforward given the use of postcode based areas in the former and grid square areas in the latter. In England and Wales, there is no guarantee that enumeration districts will have remained the same between 1971, 1981 and 1991. The same observation applies to wards, although in many areas these will have remained the same. In 1981 OPCS provided SAS data for Census Tracts (amalgamations of 1971 and 1981 EDs into larger, mostly comparable areas), which consisted of a mixture of unchanged EDs and much larger "lowest common denominator" areas. Mike Coombes' paper discusses this problem further.

It would be useful if a comprehensive account of these changes could be included in the training materials to be prepared under the Census Initiative.

6. The geography of change

Mike Coombes, University of Newcastle-upon-Tyne

There is a tension in studying change using Census data between *stability* and *interpretability*. To study such topics as the redistribution of population between North and South of the country or between urban and rural areas requires fixed boundaries for the areal units used. However, the meaning of such fixed areas is changing with change in the underlying population distribution. So, for example, one might say that Consett's hinterland has shrunk or that Leeds' built up area has grown.

Emphasis on interpretability implies that boundaries of areas will change, though often the interpretation will be revised at the same time (and altered building blocks may prevent any stability). Experience with Standard Metropolitan Areas in the USA show that the existence of stable area building blocks (counties) have a number of advantages. First, a change analysis can be carried out using data for the second time point for areas defined at the first time point. Second, the change in functional areas between the first time point and the second can be studied. Third, a check can be carried out on the initial change analysis can be carried by comparing data for the second time point's areas.

What candidates are there for stable building blocks over the 1981-91 intercensal periods?

(1) Using the overlap analysis function in a standard GIS package has not had a strong vote of confidence in the GIS community unless the target comparable areas are quite large in relation to the building blocks. So, for example, it would not be sensible to try to estimate the 1981 populations within non-comparable 1991 EDs but it would be acceptable to use 1981 EDs to estimate the 1981 population within 1991 wards, where these have changed.

(2) There is a strong argument for using building blocks based on the highest common factors between 1981 and 1991 wards, in a similar way to OPCS's use of highest common factors between local authority areas and health areas when constructing population estimates. Populations from EDs in both 1981 and 1991 could be assigned or, if necessary, apportioned to the building blocks. Both the 1981 and 1991 area hierarchies would be recognisable. Almost all building blocks would consist of a number of whole or part EDs.

However, these building blocks would not be ideal. Analysis of individual areas (unless the ward was the same in 1981 and 1991) would not be recommended; they should be studied as a set.

7. Studying changes in housing availability, 1971-81-91 and the technical challenge of setting up a change file

Daniel Dorling, University of Newcastle-upon-Tyne

7.1 *The case for looking at change from 1971 to 1991*

It is important to look at population change over the last two decades, not just the last one for three reasons. The first is that some variables were only measured in 1971 and 1991, such as hours worked. The second is that some changes may only become clear over two decades, given the slow pace of some social changes. The third is that it will be informative to compare change over three censuses such as for ethnic group, though comparisons can only be approximate because the change in the meaning of the only common question asked (country of birth).

7.2 *A new approach to integrating data from 3 Censuses*

One approach to studying change over three censuses might be to extend the 1971/1981 Census Tract files constructed by OPCS after the 1981 Census. This approach is too restrictive geographically because comparable areas would be quite large across 1971-81-91. The variables that could be compared would very restricted.

A new approach is to write a program that allows unrestricted access to all the original cells in the SAS from each census so that change can be measured by the user in a flexible way. To do this requires the use of counts from the SAS at each census not any derived ratios, the construction of a wards by cells matrix and data compression to reduce storage for the variables and to increase speed of access.

7.3 *The necessity for studying detailed definitions*

It is, of course, necessary to study the definitions provided in each Census with a great deal of care, particularly when looking at housing variables. For example, a *dwelling* is accommodation *designed* to be occupied by a single household. A *household space* is accommodation *occupied* by one household or a vacant dwelling. Dwellings were counted in 1971 and 1991 but not in 1981.

8. Digitising census geography and using it in mapping

Bob Barr, University of Manchester

8.1 *Background*

In 1981 the digitised boundary file for Census wards, commissioned by the Department of the Environment, was purchased for the academic community. The dataset had to be extensively corrected by the National Online Manpower Information System (NOMIS) team at Durham University which made available their version to Manchester Computing Centre for general distribution.

Less use has been made of the digitised ward boundaries than might have been expected on the basis of experience with the 1971 SAS data, where all digitising was locally based.

Several reasons can be suggested for this situation. The technical difficulty of using the data deterred most users. The user had to familiar with a computer mapping or GIS package at more than elementary level. There had been criticism of the value of thematic mapping as an analytical tool within geography. Users had not experimented with cartographic techniques that overcame the perceptual problems posed by maps that shaded different sized areas. The ward scale was too coarse for displaying a lot of the socioeconomic variation within cities.

Considerable discussion and consultation took place between the Census Offices, central and local governments, the academic community and the commercial sector about the value of providing digitised boundary data at the smallest geographical scale for which SAS data were to be published. The outcomes were different for each country covered by a Census Office (see Section 2.2 for details). Here attention is concentrated on the England and Wales datasets.

8.2 *Digitised ED boundaries for England and Wales*

ESRC have negotiated the purchase of this dataset from the EDLINE consortium (MVA Systematica, the London Research Centre, the Ordnance Survey and Taylor Woodrow Graphics) to academic community specifications. There are four sets of specifications as follows.

8.2.1 Segments

The specification for segment files is as follows.

The digital boundary data will consist of segments.

Each segment will be digitised once only and appear only once in each county file.

Each segment will be numbered uniquely within each county file.

County boundary segments will be identical in neighbouring coun-

ties.

Segments should follow the ward boundaries on OS master ward maps wherever possible.

Segments must meet precisely at explicitly digitised nodes, and begin and end with start node and end node co-ordinates.

Segments along coastlines must be digitised to the OS High Water Mark.

Segments are the unique boundaries between a pair of contiguous areas. Polygons are a set of segments for an area arranged in a linked sequence that form a closed loop defining the complete boundary of the area. By storing the boundary data as segments, data storage requirements are reduced by nearly 50 percent as any segment will appear as part of two polygons.

If a polygon representation had been used, there would differences in the digitising of common boundaries and "slivers" could appear in the resulting maps.

The generic format of the segment file is shown in Table 6. It should be relatively simple to write software to convert this data structure into those needed for input to standard mapping or GIS packages.

8.2.2 Polygons

The Census Initiative specification for polygon files is as follows.

Polygons are defined as a clockwise sequence of segments.

Islands must be digitised clockwise.

Holes, that is fully contained EDs must be digitised anti-clockwise.

Headers should identify each ED and the number of polygons that make it up.

A minimum bounding rectangle must be given for each defined polygon.

The format of the polygon file is specified in Table 7. This does not contain any OS co-ordinates describing the boundaries but rather contains crossreferences to the relevant segments.

8.2.3 Nodes

The specification for node files is as follows.

A separate file to be provided for each county.

The number of segments closing onto each node must be given.

All nodes should normally have three or more segments closing onto them. Only 'holes' or 'islands' can have two segment nodes.

There should be no dangling segments. That is, no node may have a single segment attached to it.

The format of the node file is defined in Table 8.

8.2.4 Exceptions

A final file should be supplied containing a list of special EDs for which no boundary data are supplied. The format of the file is spelled out in Table 9.

8.3 *Commercial use of the England and Wales ED data*

Incorporation of any of these data in a product for use outside the academic sector will involve payment of use fees, which are being negotiated with the supplier.

Table 6. The structure of each county segment file

Item	Content	Columns	Format
Lines 1-10	Descriptive text		
Line 11	Number of segments	1- 5	numeric
For each segment:			
Header Line	Segment number	1- 5	numeric
	Number of points (N)	7- 9	numeric
	Area Left	6-13	alphanumeric
	Area Right	15-22	OPCS ED codes
	Accuracy Indicator	23	1=1:1250 2=1:2500 3=1:10000
	Node Start	25-29	numeric
	Node End	31-35	numeric
N lines of:-	Easting	1- 6	numeric
	Separator	7	space
	Northing	8-15	numeric
End of segment	Terminator	1	/
Final line	###END OF SEGMENT FILE###		

Notes:
1. All numeric fields must be right justified in their fields.
2. OS coordinates must include leading 0s where necessary.
3. The Header line, the N coordinate lines and the end of segment marker must be repeated for the number of segments shown on line 11.

Table 7. The structure of each county polygon file

Item	Content	Columns	Format
Lines 1-10	Descriptive text		
Line 11	Number of polygons	1- 5	numeric
For each polygon:			
Header line	Polygon number	1- 5	unique in county
	N of segments	7- 8	in this polygon
	Polygon identifier	10-17	OPCS ED code
	N of polygons	19-20	for this ED
	Polygon of N	22-23	which polygon in this ED
	Minimum Easting	25-30	OS coordinate
	Minimum Northing	32-37	OS coordinate
	Maximum Easting	39-44	OS coordinate
	Maximum Northing	46-51	OS coordinate
N segment lines	Segment number	1- 6	negative if reversed
End of polygon	Terminator	1	/
Final line	###END OF POLYGON FILE ###		

Notes:
1. All numeric fields must be right justified in their fields.
2. OS coordinates must include leading 0s where necessary.
3. The Header line, the N coordinate lines and the end of segment marker must be repeated for the number of segments shown on line 11.

Table 8. The structure of each county node file

Item	Content	Columns	Format
Lines 1-10	Descriptive text		
Line 11	Number of nodes	1- 5	numeric
For each node:-			
	Node number	1- 5	numeric
	Easting	7-12	numeric
	Separator	13	space
	Northing	14-19	numeric
	Separator	20	space
	Segments connected	21-22	numeric
Final line	###END OF NODE FILE###		

Notes:
1. All numeric fields must be right justified in their fields.
2. OS coordinates must include leading 0s where necessary.
3. The number of single line node records appearing in the file must correspond to the number given in line 11 of the file.

Table 9. Structure of each county exception file

Item	Content	Columns	Format
Lines 1-10	Descriptive text		
Line 11	Number of exceptions	1- 5	Total in file
	Text label	7-15	"SUPPRESSED"
N lines	Suppressed ED	1- 8	alphanumeric
	Grouped ED	10-17	OPCS ED codes
Text line	Text label	1-16	"END OF SUPPRESSED"
Following line	N of exceptions	1- 5	Total in file
	Text label	7-15	"SPECIAL"
N lines	SED	1- 8	alphanumeric
	Containing ED	10-17	OPCS ED codes
Text line	Text label	1-14	"END OF SPECIAL"
Final line:	###END OF EXCEPTION FILE###		

Notes:

1. The total number of lines in the file must be 10 (text header) + (number of suppressed EDs = n) + n (one for each suppressed ED) + 1 ("END OF SUPPRESSED" marker) + 1 (number of special EDs = m) + m (one for each special ED) + 1 ("END OF SUPPRESSED" marker) + 1 File terminator flag.

9. Issues in the design and production of a Census Atlas: the example of the North West region

Alex Hirschfield, Peter Brown and John Marsden,
University of Liverpool

9.1 *NW Census Atlas: background*

The Urban Research and Policy Evaluation Regional Research Laboratory (URPERRL) has been contracted to prepare two Census atlases for the North West Information Officer Group. One Atlas will cover all the North West counties while a second will focus on Cumbria. Some 22 variables will be displayed for 1991 and 5 variables measuring change for 1981-91, with 1 variable held in reserve (see Table 10 for the list of chosen indicators). Some 6 variables will be derived from the 100% Population data, 8 from the 100% Household tables, 2 will refer to the 100% Dwellings tables, 5 to the 100% Economic Activity tables, while 1 will measure qualifications and 5 employment status of residents, derived from the 10% tables. The Atlas will consist of black and white and solid colour maps. It will be produced using the GIS package Arc/Info.

9.2 *Design and production issues*

In drawing up this specification careful thought was given to the function of the Atlas, to the targeted customers and to circulation policy. The Atlas will have a reference function for planning and research and intelligence departments in the North West and provide decision makers with the background for decision taking.

9.3 *NW Census Atlas: locational data*

Before map production can begin a set of locational data files appropriate to the mapping software to be used must be created. For this project the ward boundary datafiles provided for the academic community in connection with the 1981 Census were used to test procedures. The mapping package to be used was Arc/Info.

The first problem to be solved was the conversion of the existing files in GIMMS format to ARC files in Arc/Info format. A program developed by Martin Charlton of the Centre for Urban and Regional Development Studies at the University of Newcastle upon Tyne was used for this purpose. Arc/Info requires the prior creation of an ARC file, a LABEL POINTS file and a LABELS file, examples of which are shown in Table 11. The ARC file consists of sets of coordinate pairs describing the closed line that encloses an area. Each set begins with an identifier and finishes with an END statement. The LABEL POINTS file assigns a coordinate pair to act as the point at which the label will be attached to the area (POLYGON). The LABELS file supplies these labels. The three files are linked by the unique number assigned to the polygon, known as the USER_ID.

Other problems involved areas which had islands or mudflats

associated with them. The boundaries of the areas were simplified a little.

9.5 *NW Census Atlas: shading techniques*

Several map issues had to be confronted in designing the atlas. It was decided to keep the number of shading intervals to 4 and to use the nested means technique for allocating area values to classes. The nested means method involves dividing the distribution into two using the overall mean for the population of the whole area. The means of the lower and upper parts of the distribution are then computed. The four classes have the following ranges for each variable mapped: minimum to lower half mean, lower half mean to overall mean, overall mean to upper half mean and upper half mean to maximum. The alternative would have been to use quartiles which allocate the same number of areas to each class in ranked order.

All maps except one were to be reproduced in black and white, without a backcloth (eg OS), to keep costs down.

9.6 *Conclusions*

The Census Atlas has been designed as a simple product to reach a wide audience in the North West region, including not just local government but other planning and administrative bodies, schools, private companies and voluntary groups. Implementation can begin once the SAS/LBS datasets for Cheshire, Cumbria, Greater Manchester, Lancashire and Merseyside become available.

Table 10. NW Census Atlas: selected variables

*** Dwellings**

1. Dominant dwelling type: largest share or absolute majority by ward.

2. Dwelling surplus: total dwellings - households with residents.

*** Economic activity**

3. Residents in employment as a percentage of economically active residents.

4. Female employees working part time as a percentage of economically active persons in employment.

5. Self employed (with + without employees) as a percentage of economically active persons in employment

6. Change in self employed 1981-91.

7. Economically active females as a percentage of resident females of working age (16-59 years).

*** Qualifications**

8. Residents with degrees, professional and vocational qualifications as a percentage of all residents aged 18 and over.

*** Residents in employment**

9. Persons employed in the service sector as a percentage of resident persons in employment.

10. Change in resident population in employment in skilled manual and own account non-professional worker SEGs 1981-91.

11. Residents travelling to work by car and motorcycle as a percentage of resident persons in employment.

12. Change in resident population in employment who travel to work by car or motorcycle 1981-91.

13. Social class of head of household: largest share or absolute majority by ward.

*** Population**

14. Change in resident population 1981-91

Table 10 (Continued)

15. Persons in white ethnic group as a percentage of resident persons.

16. Persons resident in private households aged 75 and over as a percentage of all persons resident in private households.

17. Persons resident in private households with limiting long term illness as a percentage of all persons resident in private households.

18. Population density: persons resident 1991 per hectare.

19. Dependency ratio: persons in households aged 0-15 and pensionable age/persons in households aged 16-pensionable age.

*** Households**

20. Households with dependent children as a percentage of households with residents.

21. Households with one adult plus one or more dependent child(ren) as a percentage of households with residents.

22. Households with two or more economically active adults in employment and no dependent children as a percentage of households with residents.

23. Households in owner occupied accommodation as a percentage of households with residents.

24. Change in households with residents in owner occupied accommodation 1981-91.

25. Households with exclusive use of bath/shower, inside wc and with central heating as a percentage of households with residents.

26. Households without a car as a percentage of households with residents.

27. Car density: total cars per hectare.

*** Reserve variable**

* Lone pensioner households as a percentage of households with residents.

Table 11 Examples of the files needed by Arc/Info

The ARCS file

```
     1
 368789, 413440
 368828, 413481
   :       :
 368789, 413440
END
     2
 363095, 405430
 363095, 405490
   :       :

END
```

The LABEL POINTS file

```
 1, 370908, 412718
 2, 362336, 409857
 3, 374319, 412547
 :    :       :
23, 382298, 411744
 :    :       :

END
```

The LABELS file

```
 1,GO2BLAA
 2,GO2BLAB
 3,GO2BLAC
 :    :
23,GO2BMAC
 :    :

END
```

10. Census data and the analysis of the spatial variation in the socioeconomic circumstances of different ethnic groups

David Owen, University of Warwick

A full written version of this paper is available from the author at the National Ethnic Minority Data Archive, Centre for Research in Ethnic Relations, University of Warwick, Coventry CV4 7AL. Here some highlights relating to the information that will become available in the 1991 Census Local Base and Small Area Statistics are picked out.

10.1 *Ethnic data from the 1991 Census*

For the first time in a British Census a question was asked about respondent's ethnicity through a self-assessment question. The classifications to be used in the LBS are shown in Table 12 along with the fuller listing of more detailed ethnic groupings, identified in the Census Test. It illustrates how the four-fold and ten-fold ethnic classifications used in the tabulations to be presented on ethnic groups have been derived from a much larger number of possible responses to the ethnic status question, identified in the Census Test.

Table 13 lists the tabulations of variables by ethnic group which together will be released in the Local Base Statistics (LBS) and the Small Area Statistics (SAS), which together will be the primary source of data for investigating geographical variations in the circumstances of ethnic groups. The LBS data will be available for all spatial scales from counties down to electoral wars, while the SAS data will also be produced for enumeration districts. Compared to 1981, the LBS data provides a much greater degree if detail on the characteristics of the general population at the ward scale, similar to that produced in the 1981 County Reports. However, the range of topics disaggregated by ethnic group is still quite limited.

10.2 *Making comparisons with the 1981 Census*

Many researchers will wish to make comparisons between the ethnic group counts from the 1991 Census and with 1981 Census information provided on country of birth of individual (SAS Table 4) and on persons living in households headed by a head born outside the UK (SAS Table 36). Haskey (1991) has developed a method that uses Labour Force Survey (LFS) data to infer ethnicity from country of birth information. The LFS contains both variables, and thus it was possible to create a matrix of the proportion of persons of each country of origin, using the 1981 LFS. This was generated for each metropolitan county and regional remainder in Great Britain. Multiplying this matrix against the vector of countries of origin drawn from the Census data yielded an estimate of the population in each ethnic group by country of birth in 1981. This procedure was carried out for each local authority district in Great Britain, using the matrix for the region, region remainder or metropolitan county it was located in.

Table 12 OPCS Ethnic classifications used in the 1991 Census

4-way classification (SAS)	10-way classification (LBS)	Full listing (Census test)
White	White	White
		Irish
		Greek/Greek Cypriot
		Turkish/Turkish Cypriot
		Mixed White
Black groups	Black Caribbean	Black- Caribbean
		Caribbean Island
		West Indies
		Guyana
	Black African	Black-African
		Africa south of Sahara
	Black other	Black-other
		Black-British
		Black-Mixed Black/White
		Black-Mixed Other
Indian/Pakistani/ Bangladeshi	Indian	Indian
	Pakistani	Pakistani
	Bangladeshi	Bangladeshi
Chinese & Others	Chinese	Chinese
	Other-Asian	E. African Asian
		Indo-Caribbean
		Black-Indian sub-continent
		Black-other Asian
	Other-other	North Africa/Arab/Iranian
		Mixed Asian/White
		British ethnic minority (other)
		British (no indication)
		Other Mixed Black/White
		Other Mixed Asian/White
		Other Mixed-Other

Table 13 Tables containing information on ethnic group available in the 1991 Census LBS and SAS

Table No.	Level	Source	Table Topic	Categories
4	100	LBS	Medical & care establishments	4
5	100	LBS	Hotels and other establishments	4
6	100	LBS/SAS	Ethnic group by age and gender	10
9	100	LBS/SAS	Economic position and ethnic group	10/4
17	100	LBS/SAS	Ethnic group of migrants	10/4
43	100	LBS/SAS	Household composition and ethnic group	10/4
49	100	LBS/SAS	Type of accommodation by ethnic group	10/4
51	100	LBS/SAS	Country of birth and ethnic group	10/4
85	10	LBS	Ethnic group of qualified manpower	10
93	10	LBS	SEG, social class and ethnic group	10

Source: OPCS Topic Report 3.

The method needs to be improved because it produces a more dispersed population by ethnic group than is probably the case. For example, persons born in India are mostly of Indian ethnic origin, but a small variety (mainly elderly) are whites born in India in the colonial era. These two groups live in very different locations and applying an average conversion matrix to the country of birth data will produce too few persons of Indian ethnicity in areas of immigrant settlement and too many persons of Indian ethnicity in areas of retirement settlement. Further local research is indicated.

10.3 *Conclusion*

The 1991 Census provides the first opportunity to produce a comprehensive picture of the spatial distribution of the main ethnic groups in Great Britain. It represents a considerable advance on the data available from the 1971 and 1981 Censuses in terms of the detail available at the ward scale: the LBS contain some 20,000 counts, compared to the 4,500 available from the 1981 SAS. While previous data has permitted analysis of the broad location pattern of ethnic minorities, analysts of the 1991 Census will be able to investigate detailed spatial differentials in occupational composition, labour market experience and housing conditions - all of which are important for assessing the experience of ethnic minorities in different local contexts, and which will provide valuable information for organisations concerned with planning the needs of ethnic minorities.

11. Country of birth, language and social status: an analysis

John Giggs, University of Nottingham

A full version of this paper will appear in *Area* in 1992 (Giggs and Pattie 1992) and here just a few of the main points are picked out. The paper examines the composition of the population living in Wales in terms of country of birth, languages spoken and social class, using published and special tables from the 1981 Census. The analysis is to be repeated using 1991 Census data and extended to look at the spatial and social characteristics of the Welsh, Scots and Irish born in England.

There has been considerable interest in the geography of language in Wales, but rather little attention paid to the relationship of language to either country of birth or social class. Maps in the paper demonstrate that the districts of North and West Wales, long the Welsh speaking heartland, are also the areas with the greatest proportions of non-Welsh born, in-migrants principally from England. Here there is a fairly direct association between the rise in in-migrant proportions and the proportional decline in Welsh speakers, though in many Welsh speaking communities the children of in-migrants from England will learn Welsh. By contrast, the districts of South Wales house populations, less than 20% of whom can speak Welsh, but 80% or more of whom were born in Wales.

When ability to speak Welsh is crossclassified by social class, however, a somewhat different picture emerges. Welsh speakers in the service class (professional, technical and managerial workers) are concentrated in South East Wales when a measure of relative concentration (the location quotient) is used. This concentration is clearly linked to the availability of higher grade jobs, often needing a Welsh speaking qualification, in and around the capital of Wales, Cardiff. By contrast Welsh speakers in the blue collar proletariat (SEGs 8, 9, 10, 11 and 15) are geographically overrepresented in North West Wales (Gwynedd) and West Glamorgan.

The evidence presented in the paper shows that contemporary Welsh society is plural rather than homogeneous. There are in Welsh society deep internal cleavages based on ethnicity, language and class. These divisions are deeply rooted in history and have produced distinctive group identities which have geographical expression. Some scholars have interpreted trends in language, ethnicity and class as evidence of a progressive erosion of the sense of a distinctly Welsh identity among Welsh people, while others have argued that Welsh identity has survived the centralising and assimilating forces in UK culture and that Welsh culture is now thriving. The 1991 Census will provide new information with which to assess these opposing theses.

12. The SARs of the 1991 Census: a new resource for social science research

Cathy Marsh and Elizabeth Middleton, University of Manchester

A full version of the paper presented here is available in Marsh (1991). The main features of this new dataset have already been introduced in section 2. Here the main points highlighted in the seminar presentation are picked out.

12.1 *Preparation of the case for the SARs*

A careful case was put together by the ESRC Working Group on the 1991 census (Marsh *et al* 1991) for purchase of the SARs for use by the academic community. One of the real concerns that had to be met was that of disclosure of information about identifiable individuals. The conclusion of the Working Group analysis was that the risk of disclosure was extremely low, essentially very near zero.

12.2 *The datasets*

Two files are to be created from the 1991 Census returns: a 2% sample of individuals (with rather limited household information) in a flat file and a 1% sample of households in a hierarchical file. The sampling will be a major element in reducing disclosure risk. However, other measures will be taken. Obviously, the records will not contain name or exact address and the geographical location will only be coded to large local authorities (with 120,000 people or more) in the 2% individual sample and to standard regions in the 1% household sample. The precise date of birth will not be held. Very large households with more than 12 members will not be represented in the sample. Many variables will be grouped before being recorded in the dataset, including occupation, industry, qualification and country of birth.

12.3 *Purchase and distribution*

A contract has been agreed between the Census Offices, ESRC and the University of Manchester to purchase the SAR datasets and to grant ESRC and the University of Manchester sole rights to distribution. Higher education institutions must sign a licence agreement on behalf of their members, who must agree to individual licences. Individuals must acknowledge the source of the data in any publications based on the SARs, and report on those works to the Census Microdata Unit every 12 months. No data can be passed on by individuals. Commercial licences can be obtained by Higher Education institutions and outside bodies can also purchase licences. The pricing strategy will be to set fairly low prices to encourage as many users as possible to use the information. Delivery of the SARs is anticipated in April 1993.

12.4 *New forms of analysis made possible by the SARs*

It will be possible, of course, for users to produce customised crosstabulations they need from the datasets. It will also be

possible to recode variables to social classifications alternative to those made available in published tables. Or alternative means of household analysis could be carried out by using alternative headship concepts or household classifications. The other advantages of the SARs to users will be the ability to carry out non-tabular analysis, to use local hardware and software for analysis and to avoid the pitfalls of using small area statistics to infer relationships.

12.5 *Services planned by the Census Microdata Unit*

Three services are planned: (1) on-line access to the SARs at MCC with associated documentation and a help service; (2) commissioned tables for researchers who need assistance in using the datasets; and (3) customised subsets for shipping to other higher education sites.

A variety of different software packages are being evaluated for use with the SARs. These include SPSS, SAS, SIR and QUANVERT implemented on three platforms: PCs running DOS, Workstations running UNIX and MCC's Amdahl running CMS. Test data from the GHS multiplied by 50 to mimic the SARs will be used in order to assess the performance of each software package. A report evaluating the alternatives will be produced before decisions are made.

13. Using the SMS from the 1981 and 1991 censuses; accessing the SMS from MATPAC

Robin Flowerdew, University of Lancaster and
Paul Boyle, University College of Swansea

This paper reviews the experience of users with using the Special Migration Statistics (SMS) from the 1981 Census and looks forward to the provision of SMS datasets from the 1991 Census.

13.1 *The structure of the 1981 SMS*

The SMS from the 1981 Census recorded the flows of migrants between wards over the year prior to the Census. The data were based on a comparison of current address at the time of the Census with previous address one year earlier, using postcode of address allocated to wards. SMS Set 1 provided a limited amount of disaggregation of these inter-ward counts by age and gender, economic position by gender, economic position by gender and marital status by gender. The simple counts of migrants between wards, disaggregated by gender were provided in the SMS Set 2.

In principle, from SMS Set 1 it was possible to derive five types of flow by the categories listed above:

(1) within the ward
(2) from the ward to other districts
(3) from other districts to the ward
(4) from countries outside GB to the ward
(5) from origins not stated to the ward.

From SMS Set 2, it was possible to derive six types of flow

(1) to, from and within wards in England and Wales
(2) to and from Scottish wards
(3) from districts where address one year ago could only be assigned to a district
(4) from countries outside Great Britain
(5) from origin not stated.

13.2 *Problems with the 1981 SMS*

However, problems occurred which prevented much analysis of SMS Set 1. In order to preserve confidentiality, a procedure known as thresholding was applied. If a particular cell in the overall migration array contained fewer than 25 migrants, the number of migrants was added to the next higher level in the geography and a zero returned in the table. It was thus not possible to tell whether a zero in a table was a real zero or a thresholded zero. This procedure also introduced an element of double counting. Investigations have shown that for flows below inter-county level the results were not consistent with those published in the Migration Regional Reports. As a result the SMS Set 1 data were not used by the academic community, and most analysis has focussed on the simpler and unthresholded SMS Set 2.

13.3 *Problems with the access software*

A package program called MATPAC (MATrix PACkage) was developed by MVA Systematica for use by local authorities and was purchased for academic use. MATPAC enables the user to extract the subset of interest of interaction data from the very large dataset of inter-ward migration flows. No further functionality is provided.

The user must carry the MATPAC analysis in four stages.

(1) The SUBAREA stage involves defining the system of interest and instructing the package to extract the relevant L block from the full matrix.

(2) The SCAN stage involves the user in specifying the zones to be used in the output, which will normally be an aggregation of the ward building blocks.

(3) The CREATE stage then involves specification of the counts to be included and produces the subfile relating to the SUBAREA, the SCANned zones and the counts.

(4) In the final stage a REPORT is produced which is a large table of the migrant counts.

The user would then usually transfer the REPORT file back to their own institution using FTP (File Transfer Protocol).

MATPAC is thus quite a complex package to use, particularly as it runs on a platform (CMS on an Amdahl mainframe) that is unfamiliar to most potential users.

13.4 *Applications of SMS*

The SMS Set 2 data have been used in a number of studies of local migration. Boyle (1991) used inter-ward migration data extensively in his doctoral study of population movement in Hereford and Worcester (see also Flowerdew and Boyle 1992) and is currently investigating inter-county migration using SMS Set 1 data. Duley and Rees (1991) have used SMS Set 1 data as input to a microsimulation model of the Leeds population as a guide to the probabilities of out-migration from wards. Rees and Rees (1990) have used SMS Set 2 data in models for projecting ward populations for the City of Swansea. Inter-ward migration data are being used in a current projection model for West Yorkshire districts (Rees 1992). Inter-ward migration data has been aggregated to provide an inter-constituency migration matrix in an electoral study. However, in general the SMS data have been underused principally because the problems introduced by the thresholding of SMS Set 1 and because of the difficulty of use of the MATPAC access software.

13.5 *The 1991 Census Special Migration Statistics*

This account precedes the SMS User Guide and so must be regarded as preliminary in content. According to OPCS, "the full SMS will be available for Local Authority Districts, and for Wards (in England and Wales) and Postcode Sectors (in Scotland) for the less detailed counts".

13.5.1 The SMS for inter-district migration

The thresholding procedures adopted for the 1981 SMS will be dropped and a simpler suppression criterion adopted that tables must contain 10 or more migrants in each cell to be published. There will be no suppression of tables showing age and sex of migrants.

The SMS for inter-district migration will contain the following tables.
(1) All migrants by age and sex (18 ages and 2 sexes)
(2) All migrants by marital condition (3 groups)
(3) All migrants by ethnic group (4 groups)
(4) All migrants by whether resident in households by whether suffering long term limiting illness (2 by 2)
(5) Migrants aged 16+ by economic position (7 groups)
(6) Wholly moving households by tenure (3 groups, 4 in Scotland)
(7) Wholly moving households by sex and economic position of head (5 groups)
(8) Persons in wholly moving households by sex and economic position of head (5 groups).

13.5.2 The SMS for inter-ward migration

These data will be provided without suppression in the following tables.
(1) All migrants by age and sex (3 ages and 2 sexes)
(2) Wholly moving households
(3) Residents in wholly moving households.

The age classification in the first tables will be: 0-15, 16-pensionable age and pensionable age and over.

13.5.3 User defined origins and destinations

There will be an opportunity for users to define their own areas of origin and destination which could consist of a single ward or groups of wards. If the area defined has a population of 25,000 or more the standard suppression procedures for inter-district SMS will apply. If the population of the area is smaller than 25,000, the user zones will be amalgamated until the gthreshold population is achieved.

These proposals should produce much more useful data on population movements. They should, however, be coupled with equivalent improvements in ease of use of the associated MATPAC91 software.

14. Demonstration of software for accessing Census data

Keith Cole, Manchester Computing Centre
John Stillwell, University of Leeds

The final session of the workshop was devoted to demonstrations of two packages for accessing, tabulating and manipulating Small Area Statistics (and Local Base Statistics) from the 1991 Census. A test version of SASPAC91 was demonstrated by Keith Cole, while a demonstration version of C91 was put through its paces by John Stillwell.

14.1 *SASPAC91: a Small Area Statistics Package for the 1991 Census*

This package is made available to Higher Education Institutions through a licensing agreement with CHEST (Common Higher Education Software Team). HEIs can purchase a five year site licence for SASPAC91. Versions of the package have been developed for a number of mainframe platforms (including IBM CMS, DEC VAX VMS, ICL VME) and for PC platforms (386 or 486 machines) running DOS (version 3.3 upwards). UNIX versions of the package are under current development.

SASPAC enables the user to:

(1) *load* 1991 Census Small Area Statistics, Local Base Statistics, Special Workplace Statistics (Sections A & B), 1981 Census and similar small area datasets into compressed files for speedy processing;

(2) *select areas* by area identifier, values of count, or distance from a point;

(3) *manipulate* the statistics to create new variables, produce user-defined areas, rank and sort areas;

(4) *print* full tables complete with invaluable text (which is not included in the Census Offices' data files), individual counts, and profiles of areas;

(5) *output* data in various file formats suitable for other specialist software, such as mapping, statistical and graphical packages.

SASPAC is developed by MVA Systematica, and is managed by the London Research Centre on behalf of the Local Government Board. For more information contact:

Hywel Davies
SASPAC Project Officer
London Research Centre
Parliament House
81 Black Prince Road
London SE1 7SZ
Telephone 071 627 9696
Fax 071 627 9606

Keith Dugmore
MVA Systematica
MVA House
Victoria Way
Woking
Surrey GU21 1DD
Telephone 0483 728051
Fax 0483 755207

14.2 *C91: a PC based package for using and managing 1991 Census data on the desktop from Powys County Council*

C91 is a 1991 Census Data Management package for PCs developed by Jon Simmonds of Powys County Council. Both single user and network versions are available. The latter will be of interest to Higher Education Institutions, for whom a site licence providing an unlimited number of copies can be negotiated. C91 requires DOS 3.1 or higher (DR-DOS 6.00 is recommended). It will run on most 286 PCs as well 386 machines and higher. C91 will also run as a standard DOS application under Microsoft Windows 3 and in Windows 386 enhanced mode C91 can be run in a window from which the cut and paste functions are available.

The main features of C91 include:

(1) *production of all standard census tables* to screen, printer or disc file;

(2) *a new area building facility*, which enables users to assemble new areas using the Census units (Districts and/or wards and/or EDs) as building blocks;

(3) the facility to *customise standard OPCS census tables*, by, for example, removing particular columns or rows;

(4) a *table design option* that enables users to combine data from different Census tables or add other data to census information;

(5) means of calculating *derived variables* from the standard counts (such as percentages, averages, subtotals and so on);

(6) the ability to *export census data* in standard formats including ASCII fixed length, ASCII comma delimited, .WKS, .WK1 and .DBF for use with other PC software such as spreadsheets, statistical analysis packages and mapping.

C91 has been integrated, by arrangement with Graphical Data Capture with the products ED91 (digitised Census ED boundaries) and MAPINFO (a GIS package for PCs).

For further details contact:

Jon Simmonds
Powys County Council
New County Hall
Llandrindod Wells
Powys LD1 5LG
Telephone 0597 826380

References

Boyle, P. (1991) *A theoretical and empirical examination of local level migration: the case of Hereford and Worcester.* PhD thesis, Department of Geography, University of Lancaster.

Duley, C. and Rees, P.H. (1991) Incorporating migration into simulation models. Chapter 13 in Stillwell, J.C.H. and Congdon, P. (eds) *Migration models: macro and micro approaches.* Belhaven Press, London. Pp.228-261.

Flowerdew, R. and Boyle, P. (1992) Migration trends for the West Midlands: suburbanisation, counterurbanisation or rural depopulation? Chapter 9 in Stillwell, J., Rees, P. and Boden, P. (eds) *Migration processes and patterns. Volume 2. Population redistribution in the United Kingdom.* Belhaven Press, London. Pp.144-161.

Giggs, J. and Pattie, C. (1992) Croeso i Gymru - Welcome to Wales: but welcome to whose Wales? *Area*, forthcoming.

Haskey, J. (1991) The ethnic minority population resident in private households - estimates by county and metropolitan district of England and Wales. *Population Trends* 63, 22-35.

Marsh, C. (1991) The sample of anonymised records. *ESRC Data Archive Bulletin* Summer 1991, Number 48, 3-9.

Marsh C., Arber S., Wrigley N., Rhind D. & Bulmer M. (1988) The view of academic social scientists on the 1991 UK Census of Population: a report of the Economic and Social Research Council Working Group. *Environment and Planning A* 20, 851-889.

Marsh C., Skinner C., Arber S., Renhole B., Openshaw S., Hobcraft J., Lievesley D. & Walford N. (1991) The case for samples of anonymised records from the 1991 Census. *Journal of the Royal Statistical Society A* 154, 2, 305-340.

OPCS & GRO(S) (1991) *Census Newsletter No.15.* Office of Population Censuses and Surveys, London and General Register Office, Scotland, Edinburgh.

OPCS & GRO(S) (1992) *Census Newsletter No.19.* Office of Population Censuses and Surveys, London and General Register Office, Scotland, Edinburgh.

Rees, P. (1992) Projecting small area populations for local authorities. *The Regional Review* 2, 2, 16-18.

Rees, G. and Rees, P. (1990) Projection methods. Section N in the Programmer's documentation for SWIS: the Swansea ward population projection and information system. GMAP Ltd, University of Leeds, Leeds.

Rees P. (1992) The ESRC/UFC-ISC 1991 Census of Population Initia-

tive: delivering the data of the decade. *ESRC Data Archive Bulletin* No.51, forthcoming.

SSRU (1990) *The OPCS Longitudinal Study: user manual.* Social Science Research Council, City University.

Produced By
School of Geography
University of Leeds
Leeds LS2 9JT
From Whom Copies May Be Ordered