WORKING PAPER 97/9

THIRD WORKSHOP
THE 2001 CENSUS-SPECIAL DATASETS:
What Do We Want?

Edited by Philip Rees

PUBLISHED DECEMBER 1997

For further copies contact the Working Paper Secretary,
School of Geography, University of Leeds, Leeds, LS2 9JT
Telephone 0113 233 3300

THIRD WORKSHOP
THE 2001 CENSUS -SPECIAL DATASETS: WHAT DO WE WANT?

Edited by Philip Rees

Contributions by
Tony Champion (University of Newcastle upon Tyne)
Keith Cole (University of Manchester)
Angela Dale (University of Manchester)
Chris Denham (Office for National Statistics)
Brian Dodgeon (City University)
Robin Flowerdew (University of Lancaster)
Martin Frost (Kings College London)
Myles Gould (University of Portsmouth)
Phil Rees (University of Leeds)
Jillian Smith (Office for National Statistics)
Frank Thomas (General Register Office Scotland)
Judith Wright (City University)

A Report to
the Economic and Social Research Council (ESRC) and
the Joint Information Systems Committee (JISC)
by the participants
in the Third ESRC/JISC Workshop Planning for the 2001 Census
8th September/9th September 1997
held at the Royal Geographical Society, London

This report summarises the presentations made and discussions held at the Third ESRC/JISC *Workshops Planning for the 2001 Census*. The report presents views of expert census users and makes recommendations to ESRC and JISC about what kinds of special statistics and associated data from the 2001 Census should be requested from the UK Census Offices. The Workshops are supported by ESRC Award H507265031.

PUBLISHED DECEMBER 1997
(Revised March 1998)

For further copies contact the Working Paper Secretary,
School of Geography, University of Leeds, Leeds, LS2 9JT
Telephone 0113 233 3300

CONTENTS

ACKNOWLEDGEMENTS

ESRC and JISC very generously support this series of *Workshops Planning for the 2001 Census - Determining Academic Community Needs and Strategy*, under ESRC Award H507265031. The UK Census Offices have been able to accept invitations to attend the Workshops and receive the inevitable mix of brickbats and bouquets. Chris Denham (ONS), Frank Thomas (GROS) and Robert Beatty (NISRA) represented their respective UK Census Offices in a good spirit of constructive consultation. Several colleagues gave their time to chair the different sessions of the Workshops and thanks are due to Mike Coombes (University of Newcastle upon Tyne), Oliver Duke-Williams (University of Leeds), Ian Turton (University of Leeds) and Heather Joshi (City University). Li Heming (University of Leeds) assisted very ably with the processing of census users' questionnaire returns. The administrative arrangements for the Workshop were very efficiently executed by Christine Macdonald (University of Leeds) and spacious rooms and quality refreshments and meals were organised by Alison Glazebrook (Royal Geographical Society with the Institute of British Geographers).

LIST OF SPEAKERS

Tony Champion
Department of Geography, Daysh Building, University of Newcastle upon Tyne, Newcastle Upon Tyne, NE1 7RU
Tel 0191 222 6000 ext. 64, Fax 0191 261 1182, email tony.champion@newcastle.ac.uk

Keith Cole
Manchester Computing, University of Manchester, Oxford Road, Manchester, M13 9PL
Tel 0161 275 6066, Fax 0161 275 6040, email k.j.cole@mcc.ac.uk

Angela Dale
Cathie Marsh Centre for Census and Survey Research, Faculty of Economic and Social Studies, University of Manchester, Dover Street, Manchester, M13 9PL
Tel 0161 275 4876, Fax 0161 275 4722, email Angela.Dale@man.ac.uk

Chris Denham
Census Division, Office for National Statistics, Segensworth Road, Titchfield, Fareham, Hants, PO15 5RR
Tel 01329 813 720, Fax 01329 813 587

Brian Dodgeon
ONS Longitudinal Study Support Programme, City University, Northampton Square, London EC1V 0HB
Tel 0171 477 8000 ext 4128, Fax 0171 477 8583, email bd@ssru.city.ac.uk

Robin Flowerdew
Department of Geography, University of Lancaster, Lancaster, LA1 4YB
Tel 01524 593 739, Fax 01524 847099, email flowedew:lancaster.ac.uk

Martin Frost
Department of geography, Kings College London, the Strand, London, WC2R 2LS
Tel 0171 873 2612, Fax 0171 873 2287, email udfa170@bay.cc.kcl.ac.uk

Myles Gould
Education Centre, St Mary's Hospital, Milton Road, Portsmouth, P03 6AD
Tel 01705 866 940, Fax 01705 866 920, email gould@geog.port.ac.uk

Phil Rees
School of Geography, University of Leeds, Leeds, LS2 9JT
Tel 0113 233 3341, Fax 0113 233 3308, email phil@geog.leeds.ac.uk

Jillian Smith
Demography and Health Group, Office for National Statistics, 1 Drummond Gate, London, SW1V 2QQ
Tel 0171 533, Fax 0171 533 5103

Frank Thomas
General Register Office Scotland, Ladywell House, Ladywell Road, Edinburgh, EH12 7TF
Tel 0131 334 0380, Fax 0131 314 4344, email gros@gtnet.gov.uk

LIST OF PARTICIPANTS

Seraphim Alvanides
The School of Geography, University of Leeds, Leeds LS2 9JT
Tel 0113 233 3348, Fax 0113 233 3308, email s.alvanides@geog.leeds.ac.uk

Robert Beatty
Northern Ireland Statistics and Research Agency, 11-13 Bloomfield Avenue, Belfast, BT5 5HD
Tel 01232 526 907, Fax 01232 526 948, email esb.nisru.9fp@nics.gov.uk

Ann Berrington
Department of Social Statistics, University of Southampton, Southampton, S017 1BJ
Tel 01703 594 549, Fax 01703 593 846, email amb6@soton.ac.uk

Wayne Codd
Room 4200W, Office for National Statistics, Segensworth Road, Titchfield, Fareham,
Hants PO15 5RR, Tel 01329 813657, Fax 01329 813638, email wayne.codd@ons.gov.uk

Mike Coombes
CURDS (NE.RRL), University of Newcastle upon Tyne, Claremont Bridge,
Newcastle upon Tyne NE1 7RU
Tel 0191 222 8014, Fax 0191 232 9259, email mike.coombes@ncl.ac.uk

Celia Curtis
Room 4300S, Office for National Statistics, Segensworth Road, Titchfield, Fareham, Hants PO15 5RR
Tel 01329 813398, Fax 01329 813587, email celia.curtis@ons.gov.uk

Wayne Diamond
Department of Social Sciences, University of Hertfordshire, Hatfield Campus, Herts. AL10 9AB
Tel 01707 285236, Fax 01707 284799, email w.j.diamond@herts.ac.uk

Oliver Duke-Williams
School of Geography, University of Leeds, Leeds, LS2 9JT
Tel 0113 233 3309, Fax 0113 233 3308, email oliver@geog.leeds.ac.uk

Catherine Hakim
Sociology Department, London School of Economics, London WC2A 2AE
Tel 0171 955 6655, email c.hakim@lse.ac.uk

Ray Hall
Department of Geography, Queen Mary and Westfield College, Mile End Road, London E1 4NS
Tel 0171 975 5408, email r.hall@qmw.ac.uk

Bernie Hayes
Department of Sociology, Queen's University of Belfast, Belfast BT7 1NN
Tel 01232 245 133 ext3712/3, Fax 01232 320 668, email b.hayes@queens-belfast.ac.uk

Ann Hockey
Planning Policy Division, Colchester Borough Council, Lexden Road, Colchester, Essex
Tel 01206 282476, Fax 01206 282475

LIST OF PARTICIPANTS (CONTINUED)

Eileen Howes
London Research Centre, 81 Black Prince Road, London SE1 7SZ
Tel 0171 787 5657, Fax 0171 787 5606, eileen.howes@london-research.gov.uk

Heather Joshi
ONS Longitudinal Study Support Programme, City University, Northampton Square, London EC1V 0HB
Tel 0171 477 8485, Fax 0171 477 8583, email hj@ssru.city.ac.uk

Peter Lee
Centre for Urban and Regional Studies, The University of Birmingham, Birmingham B15 2TT
Tel 0121 414 3645, Fax 0121 414 3279, email p.w.lee@bham.ac.uk

Rachel Leeser
London Research Centre, 81 Black Prince Road, London SE1 7SZ
Tel 0171 787 5699, Fax 0171 787 5606, rachel.leeser@london-research.gov.uk

Li Heming
School of Geography, University of Leeds, Leeds, LS2 9JT
Tel 0113 233 3348, Fax 0113 233 3308, email pghl@geog.leeds.ac.uk

Kevin Lynch
ONS Longitudinal Study Support Team, City University, Northampton Square, London EC1V 0HB
Tel 0171 477 8000, ext4140, Fax 0171 477 8583, email kl@ssru.city.ac.uk

Richard McNally
Leukaemia Research Fund Centre, University of Leeds, 17 Springfield Mount, Leeds LS2 9NG
Tel 0113 233 3909/233 3911, Fax 0113 242 6065, email lrf6rjqm@gps.leeds.ac.uk

Elspeth McVey
Office for National Statistics, 1 Drummond Gate, London SW1 2QQ
Tel 0171 533 5187, email elspeth.mcvey@ons.gov.uk

Donald Morse
Data Library, University of Edinburgh, George Square, Edinburgh EH8 9LJ
Tel 0131 651 1241, Fax 0131 650 3308, email donald.morse@ed.ac.uk

Andy Peloe
School of Geography, University of Leeds, Leeds, LS2 9JT
Tel 0113 233 3348, Fax 0113 233 3308, email pgap@geog.leeds.ac.uk

Ian Turton
School of Geography, University of Leeds, Leeds, LS2 9JT
Tel 0113 233 3309, Fax 0113 233 3308, email ian@geog.leeds.ac.uk

Judith Wright
SSRU, City University, Northampton Square, London, EC1V OHB
Tel 0171 477 8000 ext 4140, Fax 0171 477 8583, email j.w@ssru.city.ac.uk

INTRODUCTION

Phil Rees

This document reports on the presentations and discussions of the Third ESRC/JISC Workshop Planning for the 2001 Census. This version is a first draft for preliminary comment and revision. The first draft is intended for distribution to Workshop participants for comment and correction. The second draft will be distributed more widely as input to discussion at the Fourth and Final Workshop to be held on 12/13 May at the University of Leeds. At the Final Workshop the recommendations from the First, Second and Third Workshops and the Census User Survey of Views on the 2001 Census will be debated and revised for presentation to ESRC/JISC and, for information, to the UK Census Offices.

The views expressed in the papers and discussion are those of the individuals concerned and do not represent the official positions of the Economic and Social Research Council (ESRC), the Joint Information Systems Committee (JISC), the Office for National Statistics (ONS), the General Register Office Scotland (GROS) or the Northern Ireland Statistics and Research Agency (NISRA).

However, the purpose of the Workshop was to present ideas, proposals and suggestions for the production of datasets from the 2001 Census, to be given serious consideration by both ESRC/JISC and the UK Census Offices for adoption in the lead up to the 2001 Census.

The report is divided into four main parts, preceded by this Introduction and succeeded by a Summary Part Five. Each part contains reports on the presentations made at the Workshop, on the discussions of each presentation and on the recommendations put forward in small group discussion. The Introduction provides a provisional examination of 40 returns from a Questionnaire survey seeking the academic community's *Views about the 2001 Census of Population* Part One of the Report contains the presentations and discussions relating to the Samples of Anonymised Records that might be generated from the 2001 Census. Part Two examines the consultation and review process underway for adding a 2001 Census link to the Longitudinal Study and also contains discussion of the related proposal for an on-line tabulation system for the 2001 Census. Part Three concerns the Workplace Statistics from the 2001 Census, addresses the lessons of experience with the 1991 Special Workplace Statistics and puts forward proposals for improvement. Part Four puts forward an analysis of the organisation and utility of migration statistics from the 1991 Census and makes radical proposals for their improvement in 2001. The Final Summary Part will be completed on revision of the report and will summarise the recommendations put forward by speakers and discussants, where they gained substantial support.

At the conclusion of each Part of the report, the points made in the Workshop discussion are gathered together. This discussion took place both after the main presentations and at the end of the Workshop in small groups. These groups produced a set of summary recommendations which are listed in each discussion chapter. Recommendations throughout the report are given a letter representing the Special Dataset to which they refer: S for the SARs, L for the LS, W for the Workplace Statistics and M for the Migration Statistics.

One theme running through the Workshop discussions, stressed by Census Office participants, was the need to make strong "business" cases for all the proposals being put forward by the academic community. A sister document, being prepared by Phil Rees and Angela Dale, will put forward the academic community's business case for proposed Census questions. This will use the full results of the consultation of census users' views. Without well specified relevant questions in the 2001 Census, there can be no information in the 2001 Census on a topic of interest to researchers.

CHAPTER 1
VIEWS ABOUT THE 2001 CENSUS OF POPULATION

Phil Rees

The Economic and Social Research Council together with the Joint Information Systems Committee (of the Higher Education Funding Councils) have funded the purchase of computer readable data from the 1991 Census for use in academic research and teaching. The data have been supported by two Census Programmes (1991-96 and 1996-2001), which ensure smooth delivery of the data to the individual user, free at the point of use. These Census Programmes fund four data units which maintain and develop the different data sets and provide help, documentation and training. The 1991-96 programme also involved some two dozen projects located in UK universities that undertook development work (particularly on interfaces to the data and new derived products), preparation of training materials and tutorials and research into topics using more than one of the datasets. In the 1996-2001 programme a small set of development projects are underway researching confidentiality issues in microdata and table data, researching the proper organisation of look up tables, developing web and relational database systems for accessing census data and the refinement of methods of generalising census boundaries.

If a similar programme of activity is to be funded based on the 2001 Census, then careful justification of census data purchases and support must be assembled. In order to put together the case, ESRC/JISC have funded a set of consultation activities within the academic community, of which this Workshop is an example. They have also commissioned a survey of census user opinion which was carried out during August and September 1997. The survey involved mailing sets of questionnaires to all departments in UK Universities carrying out social science research, to all other institutions recognised for receipt of grants by ESRC and to all members of the various census user lists held by the Census Programme Units. The questionnaire was advertised via the census-uk@mailbase list and made available on the http://census.ac.uk/ Web site maintained by the Census Dissemination Unit at Manchester Computing. About 1200 hundred questionnaires in all were sent out. The survey was voluntary and returns came inevitably from keen census users. It cannot be regarded as a representative sample of all census users in the academic community but, when finished, will be an important compendium of views of the most interested researchers. This Chapter describes results of the survey based on the first 40 returns. A fuller report will be based on a likely 120 responses and will be presented in January 1998 at the Annual Conference of the Royal Geographical Society (with the Institute of British Geographers) to be held at Kingston University.

The questionnaire will also yield information of use to the UK Census Offices in their preparation of the case for the 2001 Census, to be incorporated in a White Paper in late 1998. The first part of the questionnaire was therefore designed explicitly to gauge the need for particular questions and bundles of questions. The results from the full survey returns are to be used to produce the academic sector's "business" case for the census to be delivered to the UK Census Offices.

The questionnaire was organised in 4 sets of questions covering the following aspects:

1. Topics and questions proposed
2. Proposed changes in concepts for the 2001 Census
3. Proposed changes in outputs for the 2001 Census
4. The ESRC/JISC Census Programme for the 2001 Census

Some caution must be exercised in interpreting and using these preliminary results, but the shape of academic opinion about what is wanted from the 2001 Census is already clear. They want data that are comprehensive across topics, comparable over time, easy to access, accurate. They are fairly traditional in their requirements wishing to be able to use old (1991) geographies, old software (SPSS, SASPAC) and systems of access with which they are familiar, an evolution from the 1991 Census rather than a revolution.

Results have been assembled in a set of 27 tables placed at the end of the report, which can be consulted for full details. Tables 1.1 and 1.2 report on users' need for particular questions. Tables 1.3 reports on reactions to some innovatory approaches to the production of output statistics in the period after the 2001 Census. Tables 1.4 to 1.7 measure user experience with and views on area statistics. Tables 1.8 and 1.9 report experience and view on the SARs microdata, while parallel results for Migration Statistics are contained in Tables 1.10 and 1.11, for Workplace Statistics in Tables 1.12 and 1.13, for the Longitudinal Study in Tables 1.14 and 1.15. Responses about boundary data are given in Tables 1.16 to 1.18, while users report on use of and need for look up tables in Tables 1.19 and 1.20. Table 1.20 tests opinion on a proposed innovation for output delivery, an on-line tabulation service. Tables 1.22 through 1.25 provide user evaluations of the deliverables of the Census Programme units: registration, help, documentation, training, access and software interfaces. In this chapter we pick only a few highlights, leaving a full discussion to the report on the full set of returns.

1.1 Topics and questions proposed

Table 1.1 report user views on household topics proposed for the 2001 Census. All questions receive majority support when the "essential", "highly desirable" and " of interest" responses are summed. However, if we focus only on the "essential" and "highly desirable" categories then, there is not great support for the new *garden or yard* question nor for the extension of the question about *lowest floor level of accommodation*, previously used in Scotland, and there is less support for the *exclusive use of bath/shower/toilet*, probably because very few dwelling units now lack this feature. If questions are to be dropped then these are the best candidates, as there is much stronger support for all individual questions.

When Table 1.2 is examined, the new questions proposed receive very different levels of support. There is overwhelming support for the *income* question: only three other questions are regarded as essential for respondents' research (*employment status*, *main job* and *ethnic group*). There is also enthusiasm for a *general health* question and good support for the two questions on *unpaid help* and the *number of paid jobs* and *years since paid jobs* questions. There is, however, less support for the question on *number of people employed* which some respondents saw as demanding knowledge that many people would not have. Finally, among the new questions least enthusiasm was expressed for the question on religion. Only 42.9% saw this question as "essential" or "highly desirable", compared with the equivalent 94.4% who supported the income question. In addition some respondents expressed the worry that a question on religious group would prove contentious and therefore a threat to census coverage.

Respondents were invited to comment on question wording and coding. The question that drew most concern was the income question. Practically every respondent who commented called for the top category used in the June 1997 Census Test to be sub-divided to capture more detail at the upper end of the income spectrum. The banding used in the Census Test was:

No.	Per week	Per year (approximately)
(1)	Nil	Nil
(2)	Less than £60	Less than £3,000
(3)	£60 to £119	£3,000 to £5,999
(4)	£120 to £199	£6,000 to £9,999
(5)	£200 to £299	£10,000 to 14,999
(6)	£300 to £479	£15,000 to 24,999
(7)	£480 or more	£25,000 or more

One respondent had the following germane comment that summed up responses

> "This is far and away the most important new question. It will make the census far more valuable. There are two major problems with the wording. First the top band is too low (25,000) and it should at very least be set at the top rate of tax income (currently 26,100 - but will be higher in 2001). Second net income is of more use than gross - Britain is alone in Europe in asking about gross income - see 1991 EUROSTAT Harmonisation Workshop. Surely every other statistical office in Europe can't be wrong?"

1.2 Proposed changes in concepts for the 2001 Census

When asked about the proposed change in the population base for the 2001 Census to bring it into line with that used for mid-year estimates, most respondents felt it was desirable (62%) or essential (15%). Clearly, respondents, being in the academic sector, were very aware of the growing number and economic importance of the student population which will be counted at their term-time residence. However, most respondents also felt that the provision of some information on a 1991 Census population base was either essential (56%) or highly desirable (29%). These sets of views are being taken seriously by the Census Offices and the problem of comparability is being considered by the Population Base and Census Processing Working Group.

The proposal for a one number census in 2001 is new (and as yet untried) but it met with majority support, with 17% regarding as essential and 42% as highly desirable for their research. Similarly, a majority of respondents agreed with the processing and publication timetable suggested to accompany the free flow processing and one number census proposals (14% strongly agreed, 50% agreed). However, they also expressed a need, a little more strongly, to see some preliminary output within one year of the census (22% strongly agree, 59% agree). When this need has been put in consultation meetings that include local and central government sectors, the suggestion has been attacked. Perhaps, academics want to find out about the population of the country with more urgency than other users.

1.3 Proposed changes in outputs for the 2001 Census

1.3.1 Area statistics from the 2001 Census

The questionnaire asked respondents about their use of census data sets. Unsurprisingly, the most heavily used was the Small Area Statistics with 80% reporting use in Table 1.3, compared with 40 % for the Individual SAR (Table 1.8), only 7.5% for the Special Migration Statistics (Table 1.10), 15% for the Special Workplace Statistics (Table 1.12), 22.5% for the Longitudinal Study 1981-91 link (Table 1.14), 62.5% the Digital Boundary Data (Table 1.16) and 40% the ED/PC Directory (Table 1.15). The latter two datasets were undoubtedly used in conjunction with the Small Area Statistics.

There has been considerable debate about proposals for *output areas* from the 2001 Census (discussed in the First ESRC/JISC Workshop in January 1997). This debate arose from the diversity of geography for which needed census data at a fine spatial scale. During 1996 the

case was debated in the Output Working Group of the UK Census Offices for and against the provision of more than one small area statistics, using different geographical bases. Eventually, a consensus was reached between census user sectors and Census Office statisticians. The following points were included in the consensus:

- There should be a uniform geography across whole UK.
- Only one standard SAS set could be produced at the smallest scale.
- Output areas should, in principle, be separated from collection areas.
- Output areas should be based on aggregations of unit postcodes.
- Output areas might be based on aggregations of addresses, but that was unlikely.
- Output areas should fit wards current in 2001 and so aggregate to the whole hierarchy of administrative areas.
- Output areas would nest into wards and postal sectors on an exact or a best fit basis, depending on whether whole or split unit postcodes were used as the building brick..
- The demand for small area statistics for other geographies should be achieved through look up tables linking output areas to other areas after 2001.
- Look up tables linking output areas to 1991 Census areas should be provided.
- Digital boundaries should be output concurrently with the census statistics.
- The detail of output statistics should be adjusted to geographical scale.

The ESRC/JISC Questionnaire attempts to gauge the degree to which this consensus matched census user needs in one crucial respect: the *nature of the output areas* to be used with the 2001 Census. Respondents were asked to say how important the following alternative geographies were for their research and to rank them:

- Enumeration districts defined for collection purposes in 2001.
- Output areas built up from unit postcodes.
- Enumeration districts or output areas from the 1991 Census.

The first alternative would repeat what has usually occurred at previous censuses. The second alternative emerged from the discussions of the Output Working Group and associated research by Martin (1997). The third alternative preserves temporal comparability and was suggested by Dorling in the First ESRC/JISC Workshop.

There was little support for using *collection areas from the 2001 Census* as the output geography with only 18% ranking this as the best option (Table 1.5). Opinion was fairly split between the *postcode option* and *1991 Census reporting areas* with 41% and 43% placing these options first. This result shows the value of consultation. In Output Working Group discussion the second option had been endorsed with enthusiasm. However, the ranking may change when the full questionnaire results are analysed.

Respondents were asked next how useful would the particular solutions to the *time comparability* problem be for their research? There was greatest support for the creation of look up tables that linked 2001 output areas to 1991 enumeration districts and output areas. Some 72% (Table 1.6) saw this option as essential or highly desirable compared with 50% for output areas that aggregated to 1991 enumeration districts and output areas and 58% who strongly supported use of 1991 enumeration districts and output areas as the 2001 small area geography.

Asked to rate the acceptability of protection measures adopted for area statistics, the only device that received support was the *random adjustment of cell counts* in tables, used with the small area statistics in the past three censuses. Nearly 80% of respondents found this measure acceptable or highly acceptable, whereas the equivalent percentages for rounding of

cell numbers, suppression of low cell counts and record swapping were 30%, 38% and 17% respectively (Table 1.7). The latter figure is particularly important as this protection measure has been often suggested by Census Office statisticians as worth experimenting with. However, currently users express hostility to the device.

1.3.2 Samples of anonymised records from the 2001 Census

Two fifths of respondents reported using the Individual SAR while just under a third had used the Household SAR (Table 1.8). This level of usage is very pleasing and is a vote of confidence in the investment of ESRC/JISC and the work of the Census Microdata Unit. The output data set was first produced from the 1991 Census but "market penetration" is already high. The substantial support for similar 2001 SARs is evidenced in the 63% of the sample finding a 1% Household SAR in 2001 essential or highly desirable, while 78% endorsed the 2% Individual SAR (Table 1.9). Some 70% saw the proposal to produce a Individual SAR with Local Authority geography but less variable detail as essential or highly desirable for their research, and 56% gave similar support to an Individual SAR with sub-Local Authority geography but even less variable detail (Table 1.9). These results provide substantial backing for the work of justifying a repeat of the 1991 SAR in 2001 and in preparing the specifications and justification for a SAR with at least LA geography.

One respondent went even further, saying

> "A 100% SAR? Life would be a lot easier - confidentiality is <u>not</u> a problem."

It is unlikely that the Census Offices would accept such a proposal, but it serves to underline the links between samples of microdata released to users, giving users the ability to design their own tables, and the desire to produce such tables for the whole population. Such a facility could not be released to users given confidentiality concerns but the Census Offices are keen to develop a better, faster and cheaper tabulation service.

1.3.3 Migration statistics from the 2001 Census

The percentage of respondents indicating an interest in the Migration Statistics from the 1991 Census was low at 43%, but for 53% of these respondents migration data were either essential or highly desirable (Table 1.11). Rather few respondents had used migration statistics with only 3 and 2 reporting use of the Special Migration Statistics Set 1 and 2 respectively (Table 1.10). There are several reasons for these low responses: (1) interaction data are of interest to a small group of demographic specialists, (2) the interaction data needed some expertise to access and analyse, (3) the software interfaces were less familiar to users than those for the SAS or SARs, (4) the value of the data was seriously lowered by the failure to include out-migrants and infant migrants in the area statistics and the by suppression applied to the inter-district migration tables. These issues are addressed in detail in Chapter 13 of the report.

The question must be asked how ESRC/JISC can get better value for money from the Migration Statistics from the 2001 Census. One respondent pointed out that the migration data are of vital importance in monitoring the way in which populations are changing and urged that ESRC/JISC tackle the problem of the migration statistics by requesting a properly specified special data set under the provisions of the Census Act, using the SARs as a model:

> "We should define these as a Special data set to be purchased by ESRC/JISC on the same basis as the SARs. Better statistics could then be negotiated."

1.3.4 Workplace statistics from the 2001 Census

The percentage of respondents indicating an interest in the Workplace Statistics from the 1991 Census was low at 40%, but for 53% of these respondents migration data were either essential or highly desirable (Table 1.13). Rather few respondents had used workplace statistics with only 6 reporting use of the Special Workplace Statistics (Table 1.12). There are several reasons for these low responses: (1) interaction data are of interest to a small group of transport specialists, (2) the interaction data needed some expertise to access and analyse, (3) the software interfaces were less familiar to users than those for the SAS or SARs, (4) the value of the data was seriously lowered by the inaccuracy of the workplace location coding. One respondent explained this fourth issue forthrightly:

> "Clearly the coding of postcodes into EDs and wards could be more accurate. This is essential for the analysis of spatial policy. This process should be properly documented and available to researchers."

These issues are addressed in detail in Chapters 9 and 10 of the report.

1.3.5 The Longitudinal Study: inclusion of the 2001 Census

Rather more use has been made of this special data set than of the interaction datasets, with 23% of the sample reporting use of the 1981-1991 Linked Census Data (Table 1.14) and 56% of those answering the question on the importance of the Longitudinal Study for their research seeing it as essential or highly desirable for their research. The research produced by LS users is highly regarded and the value to the national debate about policy of a publication such as the recent *Health Inequalities* report (ONS 1997) is immense. However, there is the need to squeeze as much value out of the investment made by both ONS and ESRC/JISC. One pertinent suggestion was made by a respondent:

> "Again there is a problem of non-disclosure of small counts - provide safe analytical environment which gives results from modelling (e.g. multi-level modelling) but not access to original data."

This idea that statistical analysis before release helps protect census data from disclosure has been put forward as well in connection with an on-line service to the whole census database.

1.3.6 Boundary data from the 2001 Census

Nearly two thirds of respondents report use of digital boundary data for England and Wales and over a third use boundaries for Scotland (Table 1.16). These usage levels approach those of the area statistics and are clearly linked with the mapping of counts and derived statistics from the SAS and LBS. The most popular scales (Table 1.17) are ward (60% reporting use), districts (55%) followed by enumeration districts/output areas (45%) and counties/Scottish regions (40%). A remarkable 96% saw boundary data as important to their research (Table 1.18). Support for the provision of digital boundary data is evidenced by the remarks of one respondent:

> "Excellent scheme which must be continued and even expanded if possible to include digital data such as roads, urban areas, rivers etc."

1.3.7 Look up tables for the 2001 Census

Look up tables relating census small areas to other geographies were widely used by researchers with 40% reporting use of the 1991 Census ED/PC Directory for England and Wales, for example (Table 1.19). Of the 27 respondents replying to the question on

importance of look up tables to their own research, 95% saw them as essential or highly desirable. The role of look up tables was identified by one respondent who saw the

> "ability to compare 1971, 1981, 1991, 2001 census areas to as small an area as possible."

as vital in their research

The responses to the questions on area statistics, boundary data and look up tables confirm that purchase of these three kinds of data as one bundle should have the highest priority in call on ESRC/JISC funds to purchase 2001 Census data.

1.3.8 Flexible outputs from the 2001 Census

A final question in the section on *Proposed changes in outputs for the 2001 Census* gauged opinion on a service, special tables, which few academic researchers have used. No one reported use of such service in the survey. Not surprisingly, therefore, only 14 out of 40 census users provided assessments of the usefulness of such a flexible census output system. Of these, however, all expressed interest and 71% said they would find it essential or highly desirable (Table 1.21). One user saw such a system as

> "the single biggest improvement, from my point of view"

while another, reflecting experience in the past, considered such a proposal as

> "too expensive and much too slow."

A number of developments currently in train suggest that this judgement is capable of being overturned. The developments include much faster and more flexible tabulation software (GROS's use of SUPERCROSS), software for checking the safety of requests for standard area statistics for non-standard areas and methods of measuring risk in aggregate census tables (research at the University of Leeds), the proposal to use statistical analysis as data protection (mentioned earlier), automatic disclosure control software (research by Statistics Netherlands and the Istituto Nazionale di Statistica Italy and associated universities) and the adoption of new software strategies by the Census Offices. If all of these ingredients could be incorporated in a well organised system within the UK Census Offices, then there would be an opportunity to design, plan and negotiate an academic interface based on the successful experience of the Longitudinal Study Support Programme.

1.4 ESRC/JISC Census programme services

The final section of the census user questionnaire asked for opinions about ESRC/JISC Census Programme services.

1.4.1 Registration procedures

Some 89% of responding users found registration procedures for 1991 Census datasets outstanding or good, and nearly half judged them outstanding (Table 1.22). This is encouraging for the staff of Census Data Units who work hard on this task and have put a lot of effort into streamlining procedures and making all the necessary information available on-line. It was also surprising in that in the initial stages many users had criticised the number, length and complexity of the forms that needed to be completed and the necessity for institutional as well as individual registration and the need for witnessed signatures on special forms. One respondent captured the flavour of these concerns well:

> "I was originally asked to get the Vice-Chancellor to personally fill out and sign the university's registration form. This is way over the top. [Despite the registration agreement] it was completely unclear how much census data you could publish in an academic publication without first submitting it to OPCS for approval."

Proposals for further improving and streamlining registration procedures for 2001 Census data sets were outlined in the questionnaire (one stop registration of institution and individual covering all data sets from the 2001 Census and previous data sets). Respondents were asked to evaluate these proposals. Some 79% found them to be outstanding or good, agreeing with the respondent who commented

> "One-stop registration for as many of the datasets as possible has to be a good move!"

However, only 15% placed the proposals in the outstanding category. Users were more cautious in awarding high marks to a proposal still to be implemented compared with marking highly a system with which they had experience. There was, however, one radical suggestion that

> "it should be illegal to restrict access to census data."

1.4.2 Help with census data

Help provided by the Census Programme was rated highly with 15% seeing it as outstanding and 64% as good. However, there were 15% who had not used a Census Programme service, indicating that even among researchers using census data awareness needs to be raised.

1.4.3 Documentation and training

Most help with using census data is delivered through documents and training course. Table 1.25 reports on respondents' assessments of the usefulness of the different "help delivery packages". All "packages" bar one receive majority endorsement as essential or highly desirable. On-line information received most "essential" votes, followed by Census Office publications and searchable databases, although there is strong support for published books, Census programme documents and packs, newsletters and one day workshops. The only package not enthusiastically supported were Week Long Seminars, probably because only two of them were held. They were designed to reach key users who would pass on the knowledge within their institutions, so that perhaps their influence has been largely hidden.

1.4.4 Methods of access

Users vote overwhelmingly for networked based methods of access to census data (Table 1.26). File transfer over JANET is the only method endorsed as essential or highly desirable by a majority (70%) of respondents, while 81% see on-line login sessions as essential, while 58% see Web sessions as essential or highly desirable. These methods are widely embraced within the academic community, so the main message of the survey responses is to urge the Census Offices to catch up. In 2001 data sets should be transferred over the Web using file transfer methods and that access to census data, whether located within the academic community (safe data sets) or within the Census Offices (data sets in a safe environment) should be on-line.

1.4.5 Software for extraction and analysis of census data

The final table (Table 1.27) reports on the software packages that users employ in their research using census data. The tried and true Statistical Package for the Social Sciences,

first developed in the late 1960s in Chicago, is still every census user's favourite. It has been developed for proprietary and generic operating systems, but the Unix and Windows versions are probably the most widely used. The only data where it is not the first mentioned are boundary data. Statistical package developers have attempted to incorporate graphical and mapping software with their products but specialist suppliers have produced much better products. There was fuller discussion of the software needs associated with the 2001 Census in the second ESRC/JISC Workshop Planning for the 2001 Census, for which a report is being prepared.

1.4.6: Shape of the ESRC/JISC census programme after 2001

A final question in the survey asked users whether they had views about how the ESRC/JISC Census Programme should be developed after 2001. The responses were few (reflecting respondent fatigue perhaps). One respondent was defeatist

> "Without a precise date for the release of 2001 data -- how can we plan?"

Such problems should be turned through imaginative thinking into new opportunities. Another respondent argued that openness should be introduced to tendering for the programme

> "The 5 basic components are OK. But perhaps a dissemination programme should be added, in order to ensure that the best options are taken for user support and data dissemination, to which existing census programme units and others could apply."

A third respondent seemed to argue that the Rest of the South East had been neglected in census dissemination

> "The regionally-based workshops are particularly useful. Not always to get up and down to London for some people in the South-East region. More regional workshops please!"

It is probably too early to make any firm recommendations for activities that are four years or more years away. However, the difficulties with census data delivery timetables could have been avoided with more careful planning by both ESRC/JISC and the Census Offices.

1.5 Conclusions

It is premature to draw firm conclusions from only one third of the expected responses to the user questionnaire. In general, responses to the proposals under development by the Census Offices and the ESRC/JISC Census Programme are very positive. However, census users want as much continuity and comparability with earlier censuses as possible. This raises important issues for the planning of outputs from the 2001 Census. The Census is often caricatured as only being for "geographers", with the observation that other social scientists could make do with a large household survey. However; this stereotype needs extending to include "historians" as well. Census users need comparable social information on people over space and time.

1.6 Discussion

The one number census

Heather Joshi asked for clarification of Census Office plans for coding imputed data in the 2001 Census. The Longitudinal Study dataset would need to be confined to real census records so that a careful scheme for coding imputed items, imputed individuals and imputed households was needed. The point was noted for future reference by Census Office participants.

Non-response in the census user survey

Angela Dale pointed out that non-response should be clearly reported in presenting the results of the census user questionnaire. Non-response has been so reported in Chapter 1's tables.

Registration

Keith Cole expressed pleasant surprise at the positive evaluations of registration systems for 1991 Census. There was, however, still a need to streamline and improve the systems for the 2001 Census to reduce the work required of both users and Census Programme Units.

The population base

Chris Denham said that the Census Office proposals for changing the population base in the 2001 Census meant that there was no possibility of fulfilling user demands for some 1991 population base counts. *The Editor comments*. Ways of making the link between term-time residence and parental residence for the population under-18 years are being considered and might be extendible to the student population in general. Some clarification of the current Census Offices' thinking on this issue would be helpful.

The business case for questions

Chris Denham stressed the need to use the information from the census user questionnaire on views about census questions as part of the business case from the academic community. It was unfortunate that the questionnaire had not asked users to identify how questions were linked and interrelated. The business case must be framed as a matrix of questions (rows) by applications/uses (columns).

Table 1.1: Household topics proposed for the 2001 census

Q.	Topic	N	% of respondents answering question				
			Essential	Highly desirable	Of interest	Very low priority	Not of interest
	70%+ "approval"						
H9	Tenure	121	72	14	9	0	5
H8	Cars or vans available	119	57	26	11	2	4
H1	Type of accommodation	118	51	32	10	3	3
	50-<70% "approval"						
H2	Sharing of accommodation	118	35	31	24	4	6
H4	Rooms	117	40	23	21	8	9
H10	Landlord	119	35	24	28	6	8
H6	Central heating	120	25	26	31	9	9
H11	Furnished/unfurnished accommodation	118	22	29	29	9	11
H3	Exclusive use of bath/shower/toilet	118	24	26	23	13	14
	<50% "approval"						
H7	*Garden or yard (new)*	*119*	*14*	*29*	*31*	*13*	*13*
H5	*Lowest floor level of accommodation (extended)*	*119*	*11*	*22*	*27*	*19*	*22*

Notes:

1. The rows are in descending order of % essential + % highly desirable.
2. N = number of responses = base for percentages (i.e. the row % are computed using this base).
3. New or extended topics are indicated in italics

Table 1.2: Individual topics proposed for the 2001 census (%)

Q.	Topic	N	% of respondents answering question				
			Essential	Highly desirable	Of interest	Very low priority	Not of interest
	70%+ approval						
I17	Employment status	126	82	14	5	0	0
I21	Main job	127	78	15	5	1	2
I31	*Total gross income (new)*	*127*	*67*	*25*	*4*	*2*	*2*
I8	Ethnic group	126	75	9	8	4	5
I14	*Educational qualifications (extended)*	*125*	*60*	*22*	*14*	*1*	*3*
I19	*Years since paid job (new)*	*126*	*41*	*39*	*15*	*2*	*4*
I24	Hours worked in main job	125	50	24	18	4	5
I7	Country of birth	125	57	17	17	4	6
I15	Usual address one year ago	122	48	25	16	7	5
I12	Long-term limiting illness	126	55	18	18	6	4
I18	*Number of paid jobs (new)*	*124*	*33*	*39*	*21*	*2*	*5*
I5	Student/schoolchild status	124	43	29	19	7	2
I11	*General health (new)*	*126*	*39*	*32*	*18*	*5*	*7*
	50-<70% "approval"						
I29	Mode of travel to work	127	46	24	15	10	6
I22	Main things done in job	125	40	26	26	2	6
I23	Supervision/management responsibilities	124	36	30	24	5	6
I10	*Provision of unpaid help (new)*	*125*	*25*	*34*	*22*	*10*	*9*
I26	What organisation or company makes/does	123	22	36	24	11	8
I25	Organisation or company of main job	124	23	34	27	10	7
I13	*Receipt of unpaid personal help (new)*	123	22	35	24	8	11
I30	Other activities in last week	125	27	30	23	11	9
I28	Address of workplace	125	37	20	18	13	13
I27	*Number of people employed (new)*	122	22	29	29	12	8
	<50% "approval"						
I6	Term-time address	121	28	20	26	13	13
I9	*Religious group (new)*	*122*	*23*	*18*	*34*	*12*	*12*
L1	Language (Irish, Gaelic, Welsh)	125	22	14	26	17	22

Notes:

1. The rows are in descending order of % essential + % highly desirable.
2. N = number of responses = base for percentages (i.e. the row % are computed using this base).
3. New or extended topics are indicated in italics

Table 1.3: Responses to questions about proposed changes in concepts for the 2001 Census (%)

How important is this change in the Census population base for your work?					
N=100%	Essential	Desirable	Undesirable	Very undesirable	Not of interest
112	18	57	6	5	13
How important is provision of information from the 2001 Census on a 1991 population base?					
N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
117	48	30	16	3	3
How important is the one number census proposal for your research?					
N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
88	19	42	30	5	5
The processing and publication timetable suggested is a sensible plan for producing outputs from the 2001 Census.					
N=100%	Strongly agree	Agree	Disagree	Strongly disagree	Not of interest
104	16	53	16	4	11
I would like to see some preliminary outputs from the 2001 Census produced within one year of the 2001 Census.					
N=100%	Strongly agree	Agree	Disagree	Strongly disagree	Not of interest
99	26	50	14	4	6

Notes:
1. N = number of responses = base for percentages (i.e. the row % are computed using this base).

Table 1.4: Number and percent using the 1991 Census area statistics

Area Statistics	N (out of 140)	%
Dataset		
1991 Census Small Area Statistics (SAS)	97	69
1991 Census Local Base Statistics (LBS)	80	57
1991 Census Statistics Derived from SAS or LBS	51	36
Other Area Statistics (please specify)	15	11
Scale		
1991 Census SAS for EDs, OAs	83	59
1991 Census LBS for Wards/Postal Sectors	85	61
1991 Census SAS or LBS, for districts, counties	83	59
1991 Census SAS for Constituencies	10	7
Country		
1991 Census SAS or LBS, England and Wales	83	59
1991 Census SAS or LBS, Scotland	53	38
1991 Census SAS, Northern Ireland	15	11

Notes: EDs = enumeration districts; OAs = Output Areas; N = number of responses

Table 1.5: The rating and ranking of output area options (%)

Option	N=100%	Rating category (%)				
		Essential	Highly desirable	Of interest	Very low priority	Not of interest
2001 EDs	88	19	17	33	26	5
OAs built from postcodes	99	38	39	15	2	5
1991 EDs/OAs	94	33	36	22	7	1

Option	N=100%	Rank (%)		
		1	2	3
2001 EDs	69	15	30	55
OAs built from postcodes	69	58	25	17
1991 EDs/OAs	69	31	46	24

Notes:
1. N = number of responses
2. ED = enumeration district (census collection area)
3. OA = output area (census publication area)

Table 1.6: Rating of solutions to the time comparability problem (%)

Proposal	N = 100%	Essent-ial	Highly desir-able	Of interest	Very low priority	Not of interest
Option 1:						
Look up tables linking 2001 OAs and grid cells	81	14	27	41	11	7
Option 2:						
Look up tables linking 2001 OAs and 1991 EDs/OAs	80	38	41	15	4	3
Option 3:						
Common zones from 2001 OAs and 1991 EDs/OAs	78	14	36	40	6	4
Option 4:						
2001 OAs that aggregate to 1991 EDs/OAs	83	18	42	28	11	1
Option 5:						
1991 EDs/OAs as 2001 output areas	78	13	40	22	23	3

Notes:
1. N = number of responses
2. ED = enumeration district (census collection area)
3. OA = output area (census publication area)

Table 1.7: The acceptability of disclosure protection measures (%)

Proposal	N=100%	Highly acceptable	Acceptable	Indifferent	Unacceptable	Not of interest
Small random changes to cell counts (e.g. -1, 0, +1)	88	16	47	15	19	3
Rounding of cell numbers (e.g. to base 3 or base 5)	86	7	33	26	28	7
Suppression of cells with low counts	88	8	33	17	39	3
Swapping of a small number of records between areas	88	2	16	21	57	5

Notes: N = number of responses

Table 1.8: Number and percent using the 1991 Census Samples of Anonymised Records

Microdata: Samples of Anonymised Records	N	%
1991 Census 1% Household File, Great Britain or Northern Ireland	43	31
1991 Census 2% Individual File, Great Britain or Northern Ireland or UK	52	37

Notes: N = number of responses (out of 140)

Table 1.9: Rating of 2001 SAR proposals in importance for research (%)

Proposal	N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
1% Household (hierarchical) SAR	77	40	34	18	5	3
2% Individual SAR with more derived household variables	78	37	37	18	4	4
Individual SAR with LA geography but less variable detail	77	21	38	23	9	9
Individual SAR with sub-LA geography but with even less variable detail	76	15	28	26	18	13

Notes: N = number of responses (out of 140)

Table 1.10: Number and percent using the 1991 Census Migration Statistics

Migration Statistics	N	%
National Migration Tables, Great Britain (published volumes)	22	16
Regional Migration Statistics, Great Britain (computer readable tables)	10	7
Local Base Statistics, Great Britain (migration tables)	25	18
Small Area Statistics, Great Britain (migration tables)	28	20
Special Migration Statistics, Great Britain, Wards/Postal Sectors (ONS Set 1)	17	12
Census, Special Migration Statistics, Great Britain, Districts (ONS Set 2)	12	9
Census Special Migration Statistics, Great Britain, Districts (Leeds re-estimates)	5	4
Other Area Statistics	0	0

Notes: N = number of responses (out of 140).

Table 1.11: Importance of the Migration Statistics for research (%)

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
61	36	30	20	7	8

Notes: N = number of responses (out of 140)

Table 1.12: Number and percent using the 1991 Census Workplace Statistics

Workplace Statistics	N	%
Transport and Workplace National Tables (published volumes)	19	14
Local Base Statistics (transport tables)	21	15
Special Workplace Statistics, Great Britain, Wards/Postal Sectors (Set A: residence zones)	23	16
Special Workplace Statistics, Great Britain, Wards/Postal Sectors (Set B: workplace zones)	23	16
Special Workplace Statistics, Wards/Postal Sectors (Set C: flows)	23	16
Other Workplace Statistics	2	1

Notes: N = number of responses (out of 140)

Table 1.13: Importance of Workplace Statistics for research (%)

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
60	42	16	28	13	2

Notes: N = number of responses (out of 140)

Table 1.14: Number and percent using the Longitudinal Census data sets

Longitudinal Study data set	N	%
1971-1981 Linked Census Data	17	12
1981-1991 Linked Census Data	28	20
1971-1981-1991 Linked Census Data	20	14
Other LS Data (e.g. births, deaths, cancers),please specify	8	6

Notes: N = number of responses (out of 140)

Table 1.15: Importance of the Longitudinal Study for research

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
66	32	35	24	6	3

Notes: N = number of responses

Table 1.16: Number and percent using the 1991 Census Boundary data sets

Boundary data	N	%
1991 Census Digital Boundary Data, England and Wales	62	44
1991 Census Digital Boundary Data, Scotland	31	22
1981 Census Digital Boundary Data, Great Britain	25	18
Other Digital boundary data	5	4
OS Digital Maps as background	16	11
OPCS/GROS ED planning maps (microfilm or paper copies)	17	12

Notes: N = number of responses (out of 140)

Table 1.17: The scales at which boundary data were used

Scales used	N	%
Enumeration districts/Output Areas	51	36
Wards/Postal Sectors	58	41
Districts	45	32
Counties/Scottish Regions	30	21
Regions (Standard or Government Office)	19	14
Other zones	8	6

Notes: N = number of responses (out of 140)

Table 1.18: Importance of Boundary Data for research (%)

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
74	73	14	11	1	1

Notes: N = number of responses (out of 140)

Table 1.19: Number and percent using the 1991 Census Look Up Tables

Look Up Tables	N	%
1991 Census Area Master File	27	19
1991 Census ED/1991 PC Directory, England and Wales	47	34
1991 Census ED/post-1991 PC Directory, England and Wales	30	21
1991 Census Indexes, Scotland	9	6
1991 Census Ward/Functional Region, Great Britain	18	13
1991 Census Ward/Localities, Great Britain	15	11
1991 Census ED/1981 Census Ward, Great Britain	22	16
Other Look Up Tables	4	3

Notes: N = number of responses (out of 140)

Table 1.20: Importance of Look Up Tables for research (%)

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
70	46	39	9	7	0

Notes: N = number of responses (out of 140)

Table 1.21: Rating of a flexible census output system be for research (%)

N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
48	25	35	38	2	0

Notes: N = number of responses (out of 140)

Table 1.22: Satisfaction with registration procedures for 1991 Census data sets (%)

N=100%	Outstanding	Good	Satisfactory	Unsatisfactory	Service not used
35	48.6	40.0	8.6	2.9	0.0

Notes: N = number of responses

Table 1.23: Satisfaction with proposed registration procedures for 2001 Census (%)

N=100%	Outstanding	Good	Satisfactory	Unsatisfactory	Service not use1
33	15.2	63.6	21.2	0.0	0.0

Notes: N = number of responses

Table 1.24: Rating of the help provided by the Census Programme (%)

N=100%	Outstanding	Good	Satisfactory	Unsatisfactory	Service not used
33	15.2	42.4	21.2	6.1	15.2

Notes: N = number of responses

Table 1.25: Usefulness of the different forms of documentation/training (%)

Documentation forms	N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
Census Office Publications	35	51.4	28.6	11.4	2.9	5.7
Published Books	31	38.7	35.5	19.4	3.2	3.2
Census Programme Documents and Packs	33	30.3	48.5	9.1	9.1	3.0
Newsletters	33	15.2	51.5	24.2	6.1	3.0
On-line Information	33	54.5	24.2	15.2	3.0	3.0
Searchable Databases	30	40.0	33.3	3.3	13.3	10.0
One Day Workshops	33	21.2	39.4	12.1	18.2	9.1
Week Long Seminars	31	9.7	19.4	32.3	25.8	12.9

Notes: N = number of responses

Table 1.26: Usefulness of methods of access for research using the 1991 Census (%)

Methods of access	N=100%	Essential	Highly desirable	Of interest	Very low priority	Not of interest
Transfer of raw data to local computer						
DAT tape	25	4.0	4.0	8.0	16.0	68.0
Other magnetic tape	26	3.8	3.8	7.7	19.2	65.4
diskette	28	17.9	7.1	10.7	14.3	50.0
CD ROM/DVD	27	3.7	22.2	22.2	11.1	40.7
File transfer	27	59.3	11.1	7.4	3.7	18.5
On-line to server						
Login session	31	80.6	9.7	6.5	3.2	0.0
Web session	26	26.9	30.8	11.5	15.4	15.4
Customised tables (e.g. LS)	23	34.8	8.7	17.4	8.7	30.4

Notes: N = number of responses

Table 1.27: Software packages used or likely to use with census data

Dataset	Software package used and frequency mentioned		
	First	Second	Third
Area Statistics	SPSS 12	SASPAC 12	EXCEL 9
SARs	SPSS 10	EXCEL 3	USAR 1
Interaction Statistics	SPSS 2	SMSTAB 1	C PROG 1
Longitudinal Study	SPSS 7	EXCEL 2	MLN 1
Boundary Data	ARC/INFO 7	MAPINFO 6	MAPVIEWER 2

PART 1: SAMPLES OF ANONYMISED RECORDS

CHAPTER 2
WHAT DO YOU WANT FROM THE 2001 SARS?

Angela Dale

2.1 The context

By spring 1998 we need to have a clear view on the requirements for SARs from the 2001 Census. Samples of Anonymised Records were a new product from the 1991 Census, which were commissioned by the ESRC as the result of the pioneering work by Cathie Marsh, Chris Skinner and other members of the working party. Unlike some census outputs which have to be produced and laid before Parliament, the SARs are a special form of output which has to be requested and paid for by users.

We have already held a number of meetings at which users views have been sought (See Newsletter No. 7 for some early views). Further consultation will take place at the one-day conference to be held on 15th October and through the ONS subgroup of the 2001 Census Working Party.

2.2 Options for discussion

In the following paragraphs I have set out a number of different options which have arisen through the consultations which have already taken place.

I propose that we start from the concept of needing to retain, as a minimum, the approximate file structures for 1991, but then consider whether these need to be changed, or additional files requested. Whilst ONS accepted the confidentiality criteria which applied to the 1991 SARs, and there have been no recorded instances where confidentiality has been breached, they will require an assessment of the risk to confidentiality posed by the release of this level of data in the 2001 SAR. I more detail is required in 2001 SARs, we shall need to provide ONS with a full assessment of any additional risk involved. The first question, then, is whether any changes are needed to the existing files?

2.2.1 The 1 per cent Household file

It is suggested that the 1% Household file and its hierarchical structure has been of very great research value and there are strong arguments for a similar file from the 2001 Census. If the same geography is retained then comparisons across time can be facilitated. Views on any changes needed to the 1% Household file would be very welcome.

2.2.2 The 2 per cent Individual file

This file has also proved very valuable, although the discussion below highlights areas where some further thinking may be worthwhile.

Geography: There will be 49 new unitary authorities in England by May 1998, as well as UAs in Wales and Scotland. This will require re-thinking the geography of a local level file.

More household variables: There have been some requests for more information about the household in which the SAR member is located - for example more information about the household head or more variables which describe the characteristics of the household. The

effect of this would be to increase the total amount of information in the file, and therefore there are implications for disclosure risk.

Sample size: The sample size for some LAs is rather small and the grouping of local districts sometimes problematic. This has led to the suggestion of a larger sample size for the individual file - perhaps 5 per cent - although trade-offs of detail against sample size may be required. The benefits of this would be to provide much better precision at the LA area. Used at a national or regional level this would also provide a larger sample for analyses of subgroups. However, if the price paid for this was a decrease in detail, many users may be reluctant to make this trade-off.

2.2.3 A third SAR with lower-level geography

Geographers have argued the need for a SAR with a lower level geography. this could take the form of a third SAR which much less individual detail but more geographical detail - well below the 120K of the 1991 SARs. A threshold of, say, 30-40,000 would require virtually no grouping of Unitary Authorities or local districts - except Orkney, Shetland, Western Isles. Within UA/LAs, the agreed threshold would be reached by grouping wards such that re-aggregations could approximate to Health Areas as well as to UA/LAs. A similar re-aggregation might also provide a basis for some approximate comparisons between 1991 and 2001.

ONS would require that a SAR with this geographical detail would need to be offset by a large decrease in individual detail - most obviously be severely restricting occupational, industrial and educational detail - but probably also involving other variables. Careful consideration would need to be given to the sample size: the sample would need to be large enough to achieve an agreed level of accuracy at a specified level which might be larger than the areas proposed above. An additional file would obviously increase the costs of the SARs. Views on these proposals would be very welcome

2.3 Specific requirements for he next SARs

As well as the overall size and structure of the SARs, now is the time to identify particular variables where grouping has been problematic and more, or different, categories are needed. For example, it has already been suggested that the tip-coding of cars at 3 or more is too restrictive and that, in the Individual SAR, more detailed categories are needed for the number of residents in the household. If you have suggestions of this kind please send them to us, but also say WHY the extra detail is needed. To be able to make a case for more detail, it is essential to be able to argue the extra value of it.

2.4 Disclosure work to assess the case for an increase in the sample size of the SARs

We have received some additional ESRC funding for a four-month period during autumn 1997 to allow us to assess the risk to confidentiality of any changes in the size and structure of the SARs and of any increase in detail. This work is being done by Mark Elliot, with consultancy from Chris Skinner, and is starting in September.

CHAPTER 3
WHAT SORT OF SAR GEOGRAPHY DO WE NEED?

Myles Gould

Data contained in the 1991 Individual SAR are restricted to 278 geographical areas in England, Wales and Scotland that represent an amalgamation of local authority districts to form contiguous output areas with a population threshold of 120,000 (Marsh and Teague 1992; Marsh 1993). These areas were designed to guard against any disclosure of information about identifiable individuals. A decision was also made not to release microdata for other sets of spatial units (e.g. health authority areas) as this might have allowed identification where SAR areas and health authority boundaries overlapped (Marsh and Teague 1992). SAR areas therefore represent rather coarse spatial units, which are arbitrary and subject to the modifiable area unit problem (Openshaw 1984), but are the only spatial disaggregation for which census microdata are currently available. The question arises as to whether census microdata with finer and more flexible geographical output areas could be provided in the future?

3.1 How might geographers' needs be met? (some possible solutions)

If it is agreed that geographers require census microdata with fine grained geographical detail, how could this be achieved without endangering individuals' confidentiality? As Marsh and Teague (1992)note: 'In the past, those making the requests [for census microdata] had failed to reach compromise between those)often geographers) wanting fine grain area detail and those (often sociologists and demographers) wanting fine grain detail on other variables such as occupation'. This is what Rees (1997) terms the 'trade-off principle' in providing Census outputs. However, both fine geographical detail and also flexibility in the choice of variables and categorisation is required in order to develop and enhance models that include individual and spatial effects in a multilevel framework. Possible solutions to the problem of specifying outputs with finer grained geography can usefully be framed within the Marsh et al (1994) typology: safe data versus safe settings. Some thumbnail sketches of possible solutions under these headings are now discussed.

3.2 Safe data solutions: multiple SARs with different specifications

These could vary in the nature and degree of the coding of particular variables (e.g. occupation variable could be collapsed into fewer categories); whilst the number of identifiable geographical areas could be carried. These different files could all be drawn as non-overlapping samples from the Census Masterfile, although there would be cost implications. Without careful planning and consultation, this option might result in outputs that did not meet the requirements of all current and future users of census microdata and might just result in Census outputs that are little different to the SAS/LBS. Figure 1 provides a typology of existing census outputs with respect to : (1) geographical detail and (2) flexibility in the choice of variables/categories for cross-tabulation.

3.3 Solutions with data in a safe setting

An alternative solution would be to mount appropriate data analysis software on a computer that was wither controlled by the Census Offices or, more approximately, regulated and monitored by MIDAS. Such an approach is not new and has been proposed by Rhind et al (1991) who suggested that users could have access to the raw data but 'would only be permitted to receive aggregate output, suitable screened to ensure nondisclosure'. What is

proposed here is that users would have access to certain modules and routines contained in some statistical packages and multilevel modelling software (e.g. MLn, Rashbash and Woodhouse, 1996). Users would not have direct access to or sight of individual records, not would they be able to build cross-tabular output with too many variables and cells. Indeed similar arrangements already exist for analysing microdata data contained in the Longitudinal Survey. Users must submit batch jobs for approval before statistical analysis of data is processed and then the output is screened to ensure confidentiality before the results are sent back to the researcher. Alternatively, users can request a downloadable cross-tabulation of individual records with no more than seven variables and 250,000 cells for analysis on their own PC (Dale 1993). Multilevel modelling of LS data is also now possible on the ONS Amdahl computer (LS User Support Programme, 1994). However, remote access is not possible, and it is therefore not easy nor practical to undertake the type of exploratory analysis that is suited to a multilevel modelling approach (Jones 1991).

What the author proposes is that users would be given more freedom in the choice and flexibility in the specification of geographical areas, when undertaking statistical modelling and producing output that contains only statistical estimates for model parameters. Those undertaking multilevel modelling would be interested in producing estimates of variance and co-variance terms associated with geographical areas at 'higher' levels. These estimates of 'higher-level' area terms could then be graphed and mapped if appropriate software was made available in the computing environment. Users would never have direct access to individual records nor estimates of level-1 random terms associated with individuals. Hands-on multilevel modelling would be facilitated thereby enabling truly exploratory analysis of geographical and sociological variations. What is required is an appropriate and secure computer interface that would only allow the user to be presented with the necessary results and outputs of their analyses.

Such a system could possibly be modified to take account of the needs of users who wanted to undertake analysis on their office-based PC:

- Microdata could be remotely extracted from a SAR Masterfile and downloaded for fine geographical areas if the categorisation of a small number of other variables was coarsely specified in a cross-tabulation
- Microdata could be downloaded for a larger number of variables and categories (using similar criteria as those used for the dissemination of the LS) if very coarse spatial units were requested
- Such data extraction could be controlled by the submission of batch jobs that were screened before submission and shipping of data by Census Office staff or a heuristic/expert-system programme that approved user requests for data extraction
- All users' requests for data extraction and subsequent analysis could be strictly governed by user undertakings that would result in heavy fines, lifetime bans from future Census analysis, and disciplinary action if a user brought themselves and their academic institution into disrepute by attempting to breach the confidentiality rules.

It is only this flexible geographical data analysis approach that will allow a vigorous investigation of the key interpretative geographical questions that arise in census analysis. For example, one could choose a number of variables and select between different sets of spatial units and levels of analysis (e.g. individuals, wards, health districts and regions), or even design 'tailor made' spatial units. The modifiable areal unit problem could then be investigated within a multilevel modelling framework.

If the proposal described above is not possible then perhaps more academic geographers should undertake collaborative work with statisticians based at the Census Offices. Intensive empirical multilevel analysis of Census Masterfiles could have a role to play in designing appropriate geographical output areas for releasing census microdata. Work would be strictly controlled and monitored by Census Office staff, whilst a range of different and modifiable SAR output areas could be assessed. For what is needed are sets of contiguous geographical areas with boundaries that maximise within-area homogeneity whilst maximising between-area heterogeneity. Moreover, area typologies that grouped places together with similar population characteristics could be assessed, thereby facilitating the creation of functional non-contiguous SAR output areas.

References

Dale A. (1993 The OPCS Longitudinal Study. In *The 1991 Census User's Guide* (Ed Dale A. and Marsh C.), p.312 HMSO, London

Jones K. (1991) *Multilevel models for Geographical Research, Concepts and Techniques in Modern Geography* 54, Norwich: Environmental Publications.

LS User Support Programme (1994) *The OPCS Longitudinal Study: Update.* Issue Number 9, copy available from SSRU, City University, Northampton Square, London, EC1V OAR.

Marsh C (1993) The sample of anonymised records. In *The 1991 Census User's Guide* (Ed Dale A. And March S.), p.295 HMSO, London.

Marsh C, Dale A and Skinner C (1994) 'Safe data versus safe settings: access to microdata from the British Census' *International Statistical Review* 62, 35-53.

Marsh C. And Teague A. (1992) Samples of anonymised records from the 1991 Census. *Population Tends 69,* p.17.

Openshaw S. (1984) *The Modifiable Areal Unit Problem.* CATMOG 38, Geo Books, Norwich.

Rashbash J and Woodhouse G (1996) *Min Command Reference,* Institute of Education, University of London.

Rees PH (1997) 'The geography of outputs from the 2001 Census: a framework in Rees P (ed.). *The debate about the geography of the 2001 Census: Collected papers from 1995-6,* Working Paper 97/1, School of Geography, University of Leeds.

Rhind D., Cole K., Armstrong M., Chow L. And Openshaw S. (1991) *An on-line, secure and infinitely flexible database system for the national census of population,* Working Paper 14, South East Regional Research Laboratory, Birkbeck College, University of London.

CHAPTER 4
VIEWS ABOUT THE SARS AND ON-LINE ACCESS TO CENSUS INFORMATION

Chris Denham, Office for National Statistics

4.1 Census Office views on SARs and on-line access

All statements this stage are made to assist discussion and the formulation of requirements and are made without commitment by the Census Offices.

We welcome the initiative to begin work on commissioning SARs from the 2001 Census, and there is no reason in principle why they should not be produced again. The work must be user driven, and should proceed on the basis that the costs of the SARs will be recovered from the customers. The Census Offices will aim to produce early 'ball park' figures of likely costs.

4.2 Challenges in planning 2001 SARs

Changes 1991-2001

- strategic changes in the Census e.g. population base, the possible base, the possible introduction of additional imputed data as part of a one-number approach
- new Census topics and changed *categories* in existing topics
- geographical changes (UAs in Wales and Scotland; SSRs or GORs?)

The confidentiality environment - has it changed?

- The large growth in individual databases, in computing power, and in the techniques of record matching will probably necessitate a fresh risk assessment whether or not the specifications of the SARs change
- public concern may seem to have diminished, but the Census is high profile and the operation can be easily damaged if the public loses confidence
- there may be scope for new forms of record modification in the SARs to inhibit record matching, and it may be necessary for the SARs to be drawn from a modified output database if this replaces post-tabulation modification.

New requirements from customers

- generally in the direction of more detail/larger or selectively boosted samples/increased area breakdown
- changes that could increase disclosure risks should have strong business cases
- some of the requirements might be better met by other forms of output services or by having a more flexible approach to adding data to the SARs as requirements evolve.

Alternatives to SARs

- the Census Offices have adopted a strategy of having a much improved service for customised output, and have begun development work e.g. trials of SuperCROSS
- the Census Office have reviewed the possibility of providing on-line access to databases through 'confidentiality screens', but new work for 2001 (following the abortive work by Rhind, Cole and others prior to 1991) is at an early stage
- on-line access would require

  - a proven method and system

  - a clear 'sound bite' presentation of confidentiality assurances

  - a general perception that the approach did not put confidentiality at risk

  that is a favourable positive :'alignment' of all factors, as happened with the SARs in 1991.
- Some 'wild card' options like PARs (*Populations* of Anonymised Records) might also help get debate into perspective.

What option should ESRC/JISC favour?

The problem with alternatives to SARs is that costs, accessibility, and scope for more complex statistical analyses are all unknown at this stage. Whilst there might have to be an alternative solution to the "third SAR/more detailed geography" requirement, the SARs are now established and the user community has substantial control over their form and dissemination. The Census Offices would recommend continuation of the work to develop 2001 SARs and would encourage steps to ensure wider use as part of the main census dissemination process. If the Census Office launch initiatives on, say, new systems for custom output and/or on-line access to data bases, requirements developed in the SARs context could in any case be helpful to these initiatives.

CHAPTER 5
DISCUSSION AND RECOMMENDATIONS ON THE SAMPLES OF ANONYMISED RECORDS

5.1 Household relationships

Several speakers stressed the value of being able, when using census microdata, to specify the relationship of each household member to every other member. The June 1997 Census Test included a matrix relationships question which generated this information. Unfortunately, the ESRC/JISC Questionnaire failed to ask about this question, assuming that the case had been firmly established. However, some doubts about the feasibility of the full matrix question were reported to persist in the Census Offices. Several participants, however, pointed out the difficulties of creating household classifications and identifying families within households without such a question. The Census Offices were urged to persist with the full matrix form.

5.2 Geography in microdata

Heather Joshi and *Angela Dale* pointed to the existence of geographical classifications in both the LS and SARs. These are type classifications and do not link the individual or household to a particular geographical location. In the Household SAR the ONS Ward classification is attached to households. In the Individual SAR a generalised version of the GB Profiles classification of enumeration districts is to be added. In the LS the Acorn classification of the enumeration district is provided along with a set of ED characteristics. *Catherine Hakim* confirmed that sociologists were interested in geographical class but not in geographical location. *The Editor comments:* There is still a strong demand among geographers and local authority users for microdata about specific places.

5.3 Recommendations from small group discussion

Recommendation S1. *There should be harmonisation across all countries of the United Kingdom in 2001, leading to the release of UK SARs.*

Recommendation S2. *The sample sizes of the Household SAR and Individual SAR should be increased, to make possible the analysis of sub-groups.*

Recommendation S3. *It is vital that a household relationship question in matrix format be used in the 2001 Census, to improve details of families within households.*

Recommendation S4. *In the Individual SAR, the opportunity should be taken to divide areas that exceed threshold size.*

Recommendation S5. *There should speedy output of SAR products to all customer sectors.*

Recommendation S6. *There should be a Third SAR identifying smaller geographical areas with less variable detail. Its specification should, however, be discussed in the context of developments of an on-line tabulation system.*

Recommendation S7. *The SARs should include only "real" households and individuals and imputed items within these records should be flagged.*

PART 2: LONGITUDINAL STUDY

CHAPTER 6
THE CENSUS CONSULTATION PROCESS, THE LS REVIEW AND ACCESS IN A SAFE SETTING

Jillian Smith

6.1 Census consultation process

The Longitudinal Study team takes part in the census consultation process and is represented on the ONS Census Working Group. The LS team makes clear to Census Division its requirements for establishing a 2001 Census link to the Longitudinal Study. The LS team is very interested in the one number census project, in terms of using the LS as one input to coverage estimation. Work is in progress on moving from manual to automated linkage of LS census records to speed up the process and reduce the cost. Advice will be needed on the feasibility of doing this under the Census Acts of 1920 and 1990 because currently no complete computer record is kept of the names and addresses on the census returns. These are used in the linkage process.

The cost of the 2001 LS-census link is to be included in the budget for the 2001 Census. It therefore will need to be justified in the same way as census questions and output products. Support will be needed from the academic community to make this case.

6.2 Longitudinal Study review

The LS team at ONS is currently carrying out a review of past research using the LS, in collaboration with the Longitudinal Study Support Programme at City University. This review will look at data quality, the potential for adding new events in future (migrations recorded in the NHS Central Register), the Census-LS link, access issues, the issue of charging for outside use and the way in which the LS is managed.

6.3 Access in a safe setting

This will continue to be the way in which academics can access the LS. Analysis of the LS is now possible on powerful PCs running the Windows NT operating system, which makes possible use of more sophisticated statistical software such as the Multi-level Modelling Package, version 3. The normal route for academic access is via an approved project implemented through the LS Support Programme at City University. However, direct access is possible, after approval, for academics with time to spend working in the safe setting at ONS.

CHAPTER 7
ONS LONGITUDINAL STUDY TO 2001: WHAT DO WE WANT?

Brian Dodgeon

7.1 What is the Longitudinal Study (LS)?

The Longitudinal Study is a dataset which follows the lives of a sample of individuals living in England and Wales from 1971 through to the current year, recording the selected demographic events (births of new LS members, cancer registrations, immigrations, births to sample women, infant deaths, widow(er)hoods, deaths and emigrations and linking the census records of sample members at the 1981 and 1991 Censuses. The Sample is selected using four birthdays during the year and constitutes about 1.05% of the population of England and Wales (Hattersley and Creeser 1995).

The LS sample is maintained through additions of new births and immigrants but loses members through deaths and embarkations. These additions and subtractions are identified through their birthdays. The LS also contains contextual information about sample members of two kinds. The first is census information (at 1971, 1981 and 1991) on all household members. The second consists of contextual variables derived from the small area statistics for the enumeration district in which the LS member resides at each census.

7.2 What makes the LS a special dataset?

7.2.1 Unique nature of the data

The LS is the most disaggregated Census dataset there is. The variables are retained at their finest coding whereas in the SARs aggregation of variable categories has been carried out to make the data safe. The LS is protected by vetting the extracted tables before release. No perturbation of the records occurs, unlike the SAS where cell counts are randomly blurred.

The LS contains 840,000 study members. Note that only a subset of these is present at any date or throughout the observation period and that the number of members grows steadily over time through addition of newly born or newly immigrated individuals. Records of the dead and emigrants are retained in the study. The LS also contains people enumerated in the households of LS members at each census, a total of 2.5 million individuals.

Data are linked over time, so new data can be related to information known about the same individual years earlier. Vital events are linked to the census data, so that event intensities can be computed using precise person-time exposure denominators.

Users vary greatly in expertise with software (and in desire to use it) and research projects vary greatly in complexity.

7.2.2 Range of research areas and analysis

Research is carried out on a wide range of topics. These include Migration, Social Mobility, Mortality Risk, Fertility, Household Formation, Occupation/Employment, Education and Training, Health, Gerontology and Transport. Most of the research exploits the ability to

examine outcomes in each of these areas in relation to prior status and to control for a wide variety of potential explanatory variables.

Users have indicated to the LS Support Team plans to carry out a range of detailed research projects once the 2001 Census link is in place, exploiting the thirty years of event data and inter-censal transitions that will then be available, the value of the LS growing with time. Topics pencilled in include the following projects.

Topic	Research area
Fertility	The intergenerational transmission of fertility patterns
	Age-parity-duration specific fertility rates
Mortality	Mortality risk and deprivation
	Maternity and mortality
	Suicide
Health	Changes in health (disability) in relation to previous occupation, housing amenities and location
	Long-standing illness and its effects on labour force participation and living standards
Mobility	Intergenerational social mobility
	Social, occupational and geographical mobility of individuals
Employment	Youth unemployment
	Young people's experience of transitions into the labour market and in household formation
Deprivation	The relation between individual and area indicators of deprivation
Households	Household change

7.2.3 Examples of LS research

Three examples illustrate what kinds of analysis are possible using LS data.

The living arrangements of children (L. Clarke, H. Joshi and A. Brown). These projects looked at birth registration of children born in two intercensal decades, and followed them to the subsequent censuses to identify any changes in their living arrangements. The later study compared cross-sectionally the family types of children in two censuses, and identified the changes experienced by two different cohorts longitudinally. The researchers created derived variables to show if children had "lost" parents, acquired new ones, or remained with the same parents over time. The study took advantage of the special qualities of the LS by using both vital event and census data and analysed both longitudinal and cross-sectional effects. The results showed declining numbers of children stay in "intact" families.

Period parity progression ratios in England and Wales (M. Ni Bhrolchain). This project aimed to produce a set of period parity progression ratios for England and Wales for 1941-1990. The data used derived from the fertility history question in the 1971 Census, updated from registered births in subsequent years and data from the 1981 and 1991 Censuses. The data set was extended to include other people in the LS member's household. The "own child" method was used to infer maternity histories for them, as no registration data were available. The research required a massive operation to transform these variables. The resulting maternity histories were used to construct tables of numerators and denominators, and hence period life tables for each year for progression from births of each order. The results showed that fertility decline measured by rates produced in this way is not severe as when conventional rates, which take no account of birth order, are used.

Sex segregation and part-time working (L. Blackwell). This project used fertility data to look at the influence of child birth on type of occupation. Household composition and birth registration data were used to produce indicators of "maternal status" at different time points. Variables were also derived to summarise the sequence of occupation by their sex segregation, making use of double coding back to 1981 occupational categories. Results showed those who go into part-time work after childbearing (especially where the employment is female-dominated) are downwardly mobile, whereas those going back into full-time work are not (especially if the employment is male-dominated).

7.3 LS developments since 1994

7.3.1 Model 204 database

The LS is now fully established as a Model 204 database, which allows full access to all data for analysis. Variables from all sources can be combined to tailor a data set for individual research projects. Previously data could only be analysed from flat file data extracts. The Model 204 database now allows more detailed work on household structure and change. A program, *flextract*, can be used to extract datasets for use with statistical packages such as SPSS or SAS. There are also several multi-purpose datasets now available for quick response to initial queries. These contain topic-specific subsets of data already in SPSS system files or SAS datasets.

7.3.2 Revised and streamlined Idealist Data Dictionary

The Data Dictionary has now been revised and streamlined using a Windows version of Idealist with hypertext. Some 3,000 variables are fully documented and searchable. Post-1991 event variables have been included. There is refined cross-referencing between decades. The Dictionary and other documentation is now available on the World Wide Web. The URL is *http://ssru.city.ac.uk/Ls/lshomepage.html.*

7.3.3 Micro-analysis in a safe environment

Transfer of large subsets of data is now possible from the ONS IBM mainframe to an NT server in a secure environment. This increases the range of analysis software which can be employed and this now includes SAS, SPSS, MLn, GLIM, STATA and MapInfo.

7.4 What do we want?

LS Link to 2001 Census
Recommendation L1. *The LS link to the 2001 Census should be funded. Each succeeding Census adds exponentially to the value of the LS.*

LS Link to raw 2001 Census
Recommendation L2. *The link should be to raw data in the 2001 Census before imputation is done or else all imputed data should be clearly flagged.*

New 2001 Census variables
Recommendation L3. *A question should be asked about all intra-household relationships.*

Recommendation L4. *An income question should be included in the 2001 Census.*

Recommendation L5. *The educational question should capture the complete educational history, which may be achieved through multi-ticking of the proposed question.*

Issues of continuity and consistency

Recommendation L6. *The location coding of 2001 Census records should record 1991 geography as well as 2001 geography, so that a ten year migration indicator can be constructed. Fine grid reference coding of addresses would facilitate comparability between census time points (see Recommendation L10).*

Recommendation L7. *Occupation should be double coded with the 1991 classification as well as the 2001 coding. This should also apply to Social Class and the proposed new ONS Social Classification.*

Recommendation L8. *To the LS should be added derived variables summarising sequences of employment status, housing tenure, region of residence, household/family position, occupation and marital status.*

Recommendation L9. *A bank of complex derived variables from previous projects should be added to the LS.*

Recommendation L10. *Look up tables should be prepared that link the smallest output areas in the 1971, 1981, 1991 and 2001 Censuses.*

Documentation

Recommendation L11. *The LS Technical volume should be updated.*

Recommendation L12. *The LS Manual or its equivalent should be updated.*

Recommendation L13. *The LS Data Dictionary should be updated.*

Recommendation L14. *Guides should be produced on changes in classifications between censuses (particularly in occupations and social class), on the comparability of derived variables across censuses and on the complex derived variables added to the LS.*

Recommendation L15. *Tutorial material introducing the LS to potential user should be placed on the World Wide Web.*

7.5 User feedback

Consideration is being given to administering a questionnaire similar to that produced for the SARs, to ascertain views of users and potential users. At the next LS Users meeting in February 1998, there will be discussion of user needs for the 2001 Census Link.

References

Hattersley L and Creeser R (1995) *The Longitudinal Study: technical manual.* LS Studies No.15. HMSO, London.

CHAPTER 8
DISCUSSION AND RECOMMENDATIONS ON THE LONGITUDINAL STUDY

8.1 The business case for questions

Discussion returned to issue of how to frame the academic sector's business case for the 2001 Census.

Several speakers made suggestions about examples of research that could be deconstructed to reveal the census questions that yielded the census data used and that were of policy relevance. *Heather Joshi* cited research on period-parity progression ratios computable using the LS which informs the nation more reliably about long term fertility trends than cross-sectional measures. *Peter Lee* cited housing research on targeting the excluded, following a Labour Government initiative, which uses SAS/LBS data to provide the neighbourhood context for policy analysis. *Phil Rees* identified the work of Peter Batey and Peter Brown for the Dearing Report which used geodemographic classification in combination with UCAS admission data on student domicile to infer the social background of students. All such analyses use moderate numbers of indicators built from larger numbers of census counts (numerators and denominators) generated by a large subset of the questions asked in the census. The challenge in assembling the business case is to deconstruct such work and then to reassemble it in the required matrix format.

8.2 Recommendations from the small group discussion

These endorsed all of the Recommendations presented by Brian Dodgeon but added one more.

Recommendation L13. *Efforts should be made to raise awareness in Whitehall of the importance of the LS Link to the 2001 Census.*

PART 3: WORKPLACE STATISTICS

CHAPTER 9
FURTHER PERSPECTIVES ON THE SPECIAL WORKPLACE STATISTICS

Keith Cole

This paper addresses issues that have arisen as users have analysed the 1991 Census workplace and journey to work questions, which should resolved in administering equivalent questions in the 2001 Census.

9.1 1991 Questions revisited

9.1.1 The workplace address question

There were a number of problems in interpreting the 1991 Census results which should be resolved in planning the workplace question in 2001 and its coding. The 1991 Census workplace address question looked at the location of the main job. Consideration might be given to finding out the workplaces of second jobs, if survey evidence suggested that these were very significant. There were problems in the 1991 Census coding of "mainly at home" and "no fixed place", which might have led to misinterpretation. There were also problems with the responses of persons who had a weekly or monthly commute.

9.1.2 The daily journey to work question

The daily journey to work question also posed problems of interpretation. It referred by implication to the main job and to the longest part by distance (important for mode specification). These issues need to be clarified in the question instructions. Multiple ticking of mode was not allowed in 1991 because of the complex data structure that would result. The 1991 Census question did not consider travel time. It might have been useful to have had a box for home workers to tick to cross check their answers. Should these issues be reconsidered in 2001?

9.2 Accuracy and geocoding issues

9.2.1 Misallocation of postcodes in England and Wales

Errors in the Central Postcode Directory (CPD) can result in the misallocation of workplaces to census areas. In 1981 10% of workplaces were misallocated at ward level and 2% were misallocated at district level. In 1991 misallocation resulted in persons at home or with "not fixed/not stated" workplaces being incorrectly counted in flows into wards. In 1991, the imputation process used to assign large user postcodes to EDs resulted in misallocations which adversely affected particular workplace zones (e.g. Docklands and Heathrow). There is also the problem where a large user postcode spans more than one ward. This happens in the case of the University of Manchester's M13 9PL postcode. Part of the campus, east of Oxford Road falls in one ward, while the campus west of Oxford Road falls in a different ward.

9.2.2 Distance to work

Errors in the Central Postcode Directory and other factors can produce some anomalous distances. One example is that in Great Manchester 89 people were reported as walking more than 40km to work.

9.3 Comparability issues

In the outputs from the 1991 Census there were differences between the SWS and the SAS/LBS. The SWS population base excludes workers on government schemes but includes workers domiciled outside GB. There were differences in coding of mode of travel groupings (e.g. the "other" category merged with "not stated"). The grouping of SWS and SAR workplace categories differs slightly. There were also differences between the 1981 and 1991 SWS coding and geographical base. Direct comparisons below district level are, for example, not possible because of changes in ward and postcode boundaries.

9.4 Content issues

There is some duplication of information between the LBS/SAS and SWS Set A, allowing for population base differences (e.g. transport to work, cars, social class, socio-economic group and occupation). Some tables are more detailed in the LBS (e.g. mode by SEG by mode by cars) than in the SWS Set A. There is also duplication in Table 4 of the SWS Set C which provides distance information of origin/destination pairs, the distance of separation already being known.

9.5 What does the QLFS offer?

To assess what might be requested in terms of Workplace Statistics from the 2001 Census we need to assess whether there are now alternative data sources which supply many of the needs of researchers. One such alternative introduced since the 1991 Census is the Quarterly Labour Force Survey (QLFS) which began in 1992, as an expansion of the long running annual Labour Force Survey.

The QLFS is a sample of 60,000 households (5 waves) per quarter. It is carried out at county scale with a separate parallel survey in Northern Ireland. Extremely detailed information is provided on the workforce (including characteristics of the unemployed). It provides information on the journey to work including travel time but not distance. Rather limited workplace location information is provided: region of place of work, area in Greater London and whether residence and workplace is in the same Local Authority District. Information is provided at LAD scale from the Annual Local Area Database associated with the LFS from 1992 onwards. Twelve key variables are available but with broad coding and no journey to work information.

Available in the QLFS are the following variables:

- Individual characteristics (age, sex, ethnicity)
- Household characteristics (tenure, household composition, type of family unit)
- Residence details (length of time at address)
- Main job (industry, occupation, employment status, size of workplace, home working, hours, travel to work)
- Second job (industry, occupation, employment status, home working, hours)
- Seeking/not seeking work (type of employment)
- Unemployment (duration, industry last job)
- Previous employment (3 and 12 months ago)
- Education and training (qualifications)

- Health (problems, accidents)
- Income (main and second job, benefits).

Clearly, the QLFS is a very valuable source of intercensal information. What the decennial census provides that the QLFS cannot is information for sub-district areas and information on residence-workplace flows.

9.6 Thresholding and suppression

Although the 1991 SWS was based on a 10% sample of census returns, a complex set of thresholding and suppression rules were defined. For symmetrical SWS data sets (where the same geographical units were used as residence zones and workplace zones), the rules were relatively simple. The full SWS (A, B and C sets) were released if the number of employed residents in the 10% sample in each zone exceeds the SAS thresholds (50 persons and 16 households). If a zone fails SAS thresholds and has less than 5 persons working in the zone then the SWS (A, B and C) are suppressed. If a zone fails the SAS thresholds and has more than 5 persons working in the zone only the SWS B were released. The rules for asymmetrical SWS (where the geographies used for residence and workplace zones differ) were more complex. Very fortunately, the academic purchase of a ward/postcode sector SWS for all GB resulted in virtually no suppression. It is essential that the academic purchase of SWS from the 2001 Census have the same characteristics.

9.7 Recommendations for 2001

On the basis of this review of the nature of the 1991 SWS the following recommendations can be put forward for an improved set of Workplace Statistics from the 2001 Census to be purchased for academic use.

Recommendation W1. *Careful attention should be given to comparability of coding categories, population base, geographical units and content across all datasets reporting Workplace statistics from the 2001 Census.*

Recommendation W2. *Careful attention should be given to comparability of coding categories, population base, geographical units and content between Workplace statistics from the 2001 Census and the Quarterly Labour Force to enable joint use..*

Recommendation W3. *Integrate SWS Set A (and Set B where zones of residence and workplace are identical) information into the area statistics output.*

Recommendation W4. *Retain a symmetrical zoning system for residence and workplace, because this is easy for users to understand and agree upon.*

Recommendation W5. *Consider carefully whether the zones chosen for the academic purchase should be wards throughout UK, postcode sectors throughout UK, or a mixture of wards (England and Wales) and postcode sectors (Scotland) as in 1991. This consideration would only affect SWS Set C (residence-workplace flows) as it is likely that area statistics will be provided for both wards and postcode sectors throughout the UK.*

Recommendation W5. *100% coding should be used to improve the reliability of sub-district data.*

Recommendation W6. *More detailed tabulations relating to mode of travel to work should be produced, particularly for interaction data (e.g. mode of travel by SEG).*

Recommendation W7. *Origin-destination distance should be supplied as part of the interaction data to improve use (e.g. defining catchment areas).*

There are a number of issues which need to be considered by users and the Census Offices. Some of them would need a new question or modification to an existing question, for which a very strong case would need to be made.

- Do we want to identify weekly/monthly commuters?
- Do we want to identify other trip types (e.g. for school children)?
- Do we want clearer differentiation between those working at home (e.g. telecommuters) and those working with home as a base (e.g. builders)?
- Do we want improved information on those working at home and with no fixed workplace (e.g. by industry)? There were about 2.5 million people in 1991 in this position.
- Do we want information on travel time?

There are also a number of improvements needed in access to the interaction data, in techniques for mapping and visualising change over time (e.g. changes to commuting patterns). More flexible access to interaction data is needed (as discussed in the Second Workshop) but improving awareness about the research potential of interaction data is equally important.

9.8 Conclusions

It is clear that those interested in labour market analysis will increasingly need access to a variety of complementary data sources. In this context, we need to understand what the Census is attempting to measure (i.e. the characteristics of the employed population not the workforce) and what it does not measure. We also need to assess whether the Census workplace related questions are still relevant given the growing complexity of the UK labour market and changing work patterns. We also need to enhance and exploit those key features of the SWS which are not available in other datasets (e.g. level of geographical detail relating to origins and destinations, sample size and coverage). Finally, in order to promote and extend use of the interaction data, it is important that the structure, content and geographical basis of the 2001 SWS is capable of meeting the data needs of the research community.

CHAPTER 10
THE SPECIAL WORKPLACE STATISTICS - LEARNING FROM 1991

Martin Frost

10.1 The case for purchase

The dominant issue here is the need for data which provides a firm spatial link between home and jobs (contained in SWS Table C). The role of the "static" Tables A and B is being eroded by new developments such as the Labour Force Survey (LFS) and the new Annual Survey of Employment that can provide up-to-date information and , in the case of the LFS, more subtle treatment of issues like multiple jobs and unemployment surveys like the LFS, however, are still limited by their sample size so that Table B remains more or less unchallenged as a source of small area data measuring the skill/occupational structure of workplaces. This is by no means a trivial issue and it is one where I suspect that local authorities will have a strongly positive view. It can certainly help support a case that the Census should maintain a workplace question.

The role of Table C rests on the importance of *location* in the identification of both home and jobs. At the moment there appears to be two areas of work where this is critical. The first is the definition of functional areas (including Travel To Work Areas - TTWAs) where all forms of grouping algorithms need this detail. The second is in the identification of labour catchment areas. If these were approximately circular in form then reasonably detailed tabulations of travel distance by zone of destination could act as a substitute for the full O/D detail of Table C. The problem is that they are not. They are strongly "warped" by a directional effect and are dominated by inward commuting rather than outward. In general in many large cities this gives them a "rain-drop" shape with the "point" pointing towards the centre of the city and the bulk of the drop located away from the employment centre in question.

The shape and size of these non-circular areas is important to the sensible implementation of training policies and the evaluation of the impacts of regeneration efforts. They can, at best, be represented only incompletely with circular approximations.

The case for full Table C detail based on studies of energy intensity and urban form is a little weaker. Here it is often sufficient to know the distances travelled by either residents or workers without necessarily knowing the precise direction of travel. However, what is critical to these analyses is the ability to cross tabulate distance travelled by mode of transport. At the moment this cannot be achieved in A or B. Also providing full origin/destination detail allows the use of "Route finder" type software to measure "real" distance or travel time between zones as an improvement on the straight line distance used in standard tabulations.

In the past there have also been a few more sophisticated modelling exercises which have considered households' locational choices from a utility maximising perspective in which work travel has played a part. I have not seen any recent examples of this but these, too, would need the full Table C detail.

As a footnote I suspect that very little of this case will be affected by an improved SAR. Even if it is possible to introduce a workplace indicator of some sort, I doubt whether it would be sufficiently detailed to affect these types of analysis.

10.2 Improvements in access and quality

If one assumes that more accurate form filling is the job of the Census Offices then our input might start with some prior discussion with ONS on the means by which incomplete or missing postcodes of workplaces are imputed. Staff at ONS admit privately that the workplace question is the worst answered Census question and the "postcode problem" is not a trivial one. I am sure the discussion of this, and other technical issues, is better carried out before processing begins!

On the question of 100% versus sampled data I think there is no doubt from a research perspective that academic users will want the approach which triggers least suppression while providing access to the greatest detail. How this can be delivered in practice is a difficult question without prior knowledge of the trade-offs involved.

Areal specification is probably more simple. Direct comparability between 1991 and 2001 is not for many people a determining issue. Most areas are grouped in some way in order to use the data sensibly. Areas based on postcodes offer the benefit of avoiding the postcode to ward allocation that inevitably introduces errors. However, compatibility with Tables A and B and LBS/SAS bases outweighs all other considerations.

The software for 2001 access needs the following characteristics

1. Fast read/processing times (this is where Quanvert is currently good)

2. Screen access in visual form

3. The ability to group zones together easily in groupings which cut across standard ward/LA/Region definitions and are often based on the distance of zones from a specified point.

This may be facilitated by adding extra information to the labels identifying zones (e.g. grid co-ordinates of centroids). It might be possible for ONS to do this or for the ESRC version of the data to have some extra processing prior to general release. I recall that there was such a scramble to provide the 1991 data that it was not feasible to consider this.

To underpin the screen access to the information would necessitate the development of a continguity matrix for wards so that grouping could be done for continuous regions.

CHAPTER 11
REACTIONS TO PROPOSALS AND PLANS FOR CONSULTATION ON THE WORKPLACE STATISTICS TO 2001

Frank Thomas

This Chapter provides initial reactions to the proposals on Workplace Statistics on behalf of the UK Census Offices and information on the Workpackage being developed by the Census Offices covering Geographical matrices (workplace and migration). However, the views presented here do not constitute official Census Office policy and the Work Package can be changed at any time in terms of scope, exclusions and approach.

11.1 Reactions to proposals

The Census Offices are currently proceeding on the assumption that there will be a Workplace question in the 2001 Census, that the question will be subject to 100% processing and that better software will be needed in house to produce and check the interaction data before release.

The following issues raised in Chapters 9 and 10 are being addressed by the Census Offices.

- It is recognised that more information is needed about the home as a workplace, given its increasing importance.
- There should be harmonisation across data sets, particularly between the interaction statistics and the area statistics.
- Multi-ticking of some questions is being examined for feasibility in processing.
- The Census Offices are aware of the opportunity to integrate SWS Sets A and B into the area statistics (SAS/LBS equivalent).
- There is recognition that symmetrical data sets are safer, easier to produce and easier to use.
- There is recognition that there should a hierarchy in the provision of attributes of the flows related to area size (e.g. a small set of counts for fine geographies, a large set of counts for coarse geographies).
- Some tables need re-thinking in consultation with user sectors.

There are also a number of difficult issues not raised in Chapters 9 and 10 which the Census Offices will need to address.

- Coding of the workplace address is often difficult. Should an imputed address be attempted?
- In developing a one number census, how can the workplace be invented for missing households and individuals?

- Measures will be needed to protect confidentiality. The Census Offices wish to avoid excessive suppression. Work to explore the costs and benefits of alternatives such as record swapping, random adjustment of table counts, or sampling is needed.

11.2 Work package plan for geographical matrices (workplace and migration)

This section of the Chapter reproduces relevant extracts from Version 1.0 of Work Package W6200E of the UK Census Offices. This is a document which sets out the plans of the Census Offices for producing interaction statistics from the 2001 Census. Paragraph numbering is used following Census Office style. The abbreviation WP refers to Work Package.

11.2.1 Organisation

Overall project manager for Output Policy and Dissemination is Chris Denham (ONS). Work Package Manager is Frank Thomas (GROS). The Work Package Team includes Ian Máté (GROS), Philip Street (GROS) and Celia Curtis (ONS).

11.2.2 Aims/objectives

2.1 The aim is to produce a specification accepted by users following consultation for the "matrix products" from the 2001 Census. The equivalent products from 1991 were the Special Migration Statistics Sets 1 and 2 and the Special Workplace Statistics Set C. These products contain two geographies: one for the origin of a move or journey to work and one for the destination. The aim is to specify the output for each type of pair of zones for origin and destination.

2.2 The specification must take account of improvements in technology and any simplification of the concepts in order to produce a simpler and cheaper product.

2.3 The specification must take account of any third party software that my be produced for users to analyse the products. Work may be done within the work package to assist the development of such software. The software should range from routines to aggregate the statistics for flows between zones to graphics and the geographical presentation of the data.

2.4 The work must take advantage of any opportunities afforded by the creation of a national database at a stroke, 100% coding of workplace and other variables and any other developments.

11.2.3 Work package scope

3.1 The work includes the development of a specification of the output (statistical tables and other) to be produced for each pair of origin and destination zones. There may be a hierarchy of zones with increasingly detailed outputs produced a zones are made larger.

3.2 The drafting of a prospectus is a part of the work of specification.

3.3 The specification of the zones is also part of the work.

3.4 The WP will need to concern itself with the coding and imputation of remote addresses from the public.

3.5 The WP will consider specifying corresponding hierarchies of zone pairs and outputs. In 1991, the SMS Sets 1 and 2 formed such a hierarchy.

3.6 The WP will consider the relationship between the zones, the output created for pairs of origin and destination zones in order to recommend what additional measures are needed to prevent the disclosure of information about identifiable individuals or households. Some measures may have been taken before the products are produced.

3.7 Consultation with users is included.

3.8 Consultation with possible software suppliers is included.

3.9 The WP may include specification of software for the analysis of the products. Such software may be developed within or outside the Census Offices.

3.10 The WP may include specification of any system for users to specify aggregation of zones and Customer Services to deliver the resulting product.

11.2.4 Exclusions

4.1 The WP will take a lead in the Census Offices' approach to Workplace and Migration output generally for 2001 and will seek to initiate related work in other projects where appropriate.

4.2 The specification of any output counting workers by a single zone (i.e. residence only or workplace only), or counting migrants by a single zone (i.e. residence at time of Census only or former residence only) will be done in Work Package W6200D, Product specification. Note that the ONC or other process that produces a national database at a stroke provides an opportunity to include, say, workplace populations and numbers of out-migrants in "standard" SAS-like output.

4.3 Development of the output is part of W6100O, Output production.

4.4 The work will also exclude the development of any software for the analysis of the products. Such software may be developed within or outside the Census Offices.

4.5 Excluded is the acceptance before release of the final product, part of the project to which Table Acceptance will belong.

4.6 Excluded (though linked) is the specification of any topic reports on workplace or migration. The work will need to ensure consistency between SWS/SMS and related topic output in definitions used but allow differences where the medium of production restricts or, alternatively, enhances the product.

11.2.5 Approaches

5.1 A review of the specification, production and use of the 1991 products will be carried out. The review will make recommendation as to what changes from 1991 are needed for 2001.

5.2 The review will have to consider what the impact will be on SWS of coding workplace and other variables at 100%.

5.3 Development of specification of 2001 products will be done as blank tables and, if possible, by means of prototypes using 1991 data or synthetic data created for the purpose.

5.4 The WP will collaborate with any third party developers of software for the analysis and display of the statistics.

5.5 Users of the 1991 products and potential users of the 2001 products will be consulted. There may be a need to suggest (or respond to) radical as well conventional (minimum change from 1991) options. They will be asked to contribute to the review, to state requirements for 2001, to make suggestions about how the requirements may be met, to comment on proposals and any prototypes from the Census Offices, and on any software for the analysis and display of the statistics. A sub-group of the Output Working Group has been formed to become the nucleus of the consultation.

CHAPTER 12
DISCUSSION AND RECOMMENDATIONS ON THE WORKPLACE STATISTICS

12.1 Discussion

Eileen Howes (London Research Centre) confirmed that Local Authorities will need ward level data on the journey to work from the 2001 Census. The SWS is the *only* source of such local statistics. She endorsed the need to achieve harmonised coding and population bases across national, local and special workplace statistics, but expressed concern that multi-ticking of the mode question would lead to incomparability with earlier censuses.

Robin Flowerdew (University of Lancaster) urged that registration for the interaction statistics should be rolled into one general registration to encourage use. He also stressed that much easier to use software was needed to increase use of the data beyond interaction specialists.

Heather Joshi (City University) commented that it would be very useful to include TTWA codes into the LS in a suitable way. This would make possible all sorts of analyses of commuting.

Mike Coombes (University of Newcastle upon Tyne) discussed the choice of zone specification for the academic purchase of SWS Set C in 2001. On balance he preferred use of wards throughout the UK to selection of postcode sectors, for better comparability with 1991 statistics and because wards had above threshold residential populations thus minimising the need for modification and , above all, avoiding suppression on 1991 rules. He stressed four key points: (1) the small are data are essential because small area data can be aggregated to non-standard areas; (2) this flexible aggregation requirement rules out the use of suppression and rounding; (3) a form of Barnardisation may be acceptable but seems likely to be problematic with a dataset in which the vast majority of cells have values of 1 or zero; and (4) thus sampling may be the most acceptable option, but the problems of a small sample are very evident in the current dataset so there should be pressure to code the maximum size of sample which can be justified against confidentiality criteria (and past practice of taking a small sample simply to cut costs should be resisted). A final point he made stressed that if the One Number Census project involved imputing workplaces - somehow - then these cases should be flagged as such.

12.2 Recommendations from the small group discussion on the Workplace Statistics

Recommendation W8. *The SWS Sets A and B should be included in the Area Statistics (subject to comparability with SWS Set C and 1991 SWS).*

Recommendation W9. *The area specification chosen for the SWS Set C academic purchase should be the same as that used in the Area Statistics.*

Recommendation W10. *Because of the trade off between table details and modification/suppression there should be a range of output specifications adjusted to spatial scale.*

Recommendation W11. *Users should be involved closely in discussions on modification/suppression, and on imputation of workplace of missing persons.*

Recommendation W12. *Detailed metadata on coding practice, modification/suppression and imputation should be provided.*

PART 4: MIGRATION STATISTICS

CHAPTER 13
MIGRATION STATISTICS FROM THE 2001 CENSUS: WHAT DO WE WANT?

Phil Rees

13.1 Measuring migration and using migration statistics

This chapter describes the statistical information on migration produced from previous British censuses, evaluates the information against analytical needs and makes recommendations about the migration statistics to be generated from the 2001 Census. In the first section conceptual and measurement issues are considered. The remaining sections of the chapter look at the form and content of the different census data sets offering migration information.

13.1.1 What migration does a census measure?

Migration is the process of relocation of people over time and space. This demographic movement alters both sending and receiving communities. In low fertility countries where the balance between births and deaths is small, migration flows are the most important determinant of population change. This is the situation in the UK.

Censuses are used to measure migration through asking a question or questions about a person's prior location, at a fixed time in the past or at the time of the last migration. For most purposes the fixed internal question is preferable. The key advantages of a census for measuring migration are comprehensiveness and the ability to generate statistics about migrant characteristics. Censuses, however, have well known disadvantages as instruments for measuring migration. UK censuses only measure migration in one year out of every ten. Multiple migrations within the year prior to the census are missed as are migrants who die in the year. Censuses also fail to measure the number of emigrants. There is little that can be done about these deficiencies in the census. Other sources of migration information must be tapped (registers, surveys).

However, there are some improvements to the way migration is measured in a census that could be made. There a migration flow that can be measured, which is not. Children under one are not assigned a migrant status. However, they have a starting address in the year (place of usual residence at time of birth) and a usual address at the time of the census. Counts of such infant migrants are an essential input to subnational population estimation and projection. The American, Australian and Canadian censuses do manage to measure these migrants by tabulating place of birth against residence at the time of the census.

Recommendation M1. *Consideration should be given again to extending the migration question to capture* ***infant migrants****, as recommended by the Migration Question Sub-Group. If a question revision is ruled out, migration tables should report the migration status of the mother (or other parent/guardian if no mother is resident in the household) for children under one.*

There is also little information captured about the status of the migrant at the start of the year at the origin location, apart from those characteristics which can be easily inferred (e.g. age one year ago) or which do not change (e.g. country of birth). For example, the important

relationship between employment status and migration cannot properly be measured because only status after the migration is known.

Recommendation M2. *Consideration should be given again to expanding the relevant questions to capture **economic position one year ago**, as recommended by the Migration Question Sub-Group.*

13.1.2 Undercounting of migration

Underenumeration is a particular problem for migration statistics. In the 1991 Census of Great Britain, the quality check (ONS, Heady, Smith and Avery 1996, Table 2.13, p.14) estimated that 0.9% more persons migrated in the year before the Census than were recorded therein. The mobility rate was 10.4% in the quality check sample but 9.5% in the census. An additional problem occurs when tables of migration flows are created. Many residents report that they had a different address one year ago but fail to provide details. The origin not stated category comprised 6% of migrants in the 1991 Great Britain Census. Migrant flows might therefore need boosting by an average of 16% to obtain a true picture of population movement. In most migration tables the number of migrants with an origin not stated are reported so that users can make their own adjustments. However, under the plans for a one number census (Jones 1997; ONS, GROS and NISRA 1997, pp.1-12) consideration should be given to imputing the missing origin information.

Recommendation M3. *Consideration should be given to the **problem of undercounting specific to the migration statistics**, because respondents systematically failed to report a different address one year before the census and because respondents who did report a different address failed to report origin details.*

13.1.3 What do researchers want to know about migration?

Accurate measures of the migration process are needed for several important purposes, which include population estimation, population forecasting, household forecasting and housing planning. A great many other activities and analyses are dependent on accurate achievements of these goals to which accurate measurement of migration contributes. Researchers also need data to examine the characteristics of migrants and to understand how characteristics affect the rate of migration. Flow data are vital for investigation of the relationship between migration and area characteristics and between migration and the friction of distance. Migration data linked to socio-economic data are important for understanding the influence of location and careers. Migration data from the Census are important for investigating the relationship between labour migration and labour market conditions. A comprehensive survey of questions about migration posed by researchers is provided in Champion and Fielding (1992).

Two kinds of migration are available from censuses to answer these kinds of questions. The first kind is made up of statistics that focus on areas, which include migrations into and out of those areas. The second kind consists of statistics that can be attached to flows between areas. Using the second kind of data the researcher can examine migration from origins to destinations. Using both kinds of migration the net balance of (intranational) migration for areas can be measured, while using flow data the net balance between specific places or types of places can be computed.

Researchers will want to classify migrants using as many of the attributes measured in the census as possible. For example, age and sex are frequently used in the demographic applications of migration data. Economic position, occupation and industry of migrant are

used in economic analyses of migration. Migration by ethnicity has been examined in detail (Champion 1996; Rees and Duke-Williams 1995c). Housing market analyses use migration data classified by household attributes.

The analysis of migration by household structure is relatively undeveloped. The UK Censuses report the migration of wholly moving households and of households where the head has migrated, together with the residents of such households. However, there are other types of migration: individual migrants moving into households with non-migrant members and migrants moving between, into or out of communal establishments.

There is also the important migration flow of students from parental homes to university/college residences and from university/college residences to first career residences on graduation. This issue has been exposed by the Migration Question Sub-Group and the decision to record students as resident at their term time address will make possible the recording of these important, hitherto undercounted flows.

Recommendation M4. *A careful study should be undertaken of ways it will be possible to classify migrants by* ***household status*** *from the 2001 Census, with a view to recommending the kinds of new tables that might be produced. This study should involve collaboration between the Census Offices and customer sectors.*

Recommendation M5. *A careful study should be undertaken of ways it will be possible to classify migrants by* ***student status*** *from the 2001 Census, with a view to recommending the kinds of new tables that might be produced. This study should involve collaboration between the Census Offices and customer sectors.*

13.1.4 Published migration statistics from censuses

In the main body of the chapter we review the statistics on migration produced in recent censuses. Some six different data sets from the UK census provide information about migration:

- National Migration Statistics
- Regional Migration Statistics
- Local Base Statistics and Small Area Statistics
- Special Migration Statistics
- Samples of Anonymised Records
- Longitudinal Study

These provide a wealth of information about migration in the UK, principally in the year before the Census. National Migration Statistics have been produced from the 1961, 1971, 1981 and 1991 Censuses and Regional Migration Statistics were published from the 1971, 1981 and 1991 Censuses. Some migration data are available in the Small Area Statistics from the 1971, 1981 and 1991 Censuses. Detailed flow data are provided in the Special Migration Statistics from the 1981 and 1991 Censuses. The Samples of Anonymised Records provided microdata from the 1991 Census, from which users can produce their own tables of migrants. The Longitudinal Study (for England and Wales) provides information about migration over 25 years by recording migrant locations at seven points in time (1966, 1970, 1971, 1980, 1981, 1990 and 1991). These last two microdata sets are not considered directly in this chapter but their shape in 2001 is being currently debated (Dale 1997; Gould 1997; Turton 1997).

13.1.5 Organisation of the chapter

The rest of the chapter is organised into five sections, which consider the main migration data sets in turn. The second section reviews provision of national migration statistics. The third section examines the regional migration statistics published from the census. The fourth section looks at migration data in the local base and small area statistics. The fifth section discusses the provision of flow data at local and small area scales in the Special Migration Statistics and tackles the confidentiality issue which is regarded as acute for such origin/destination flow data. The final section collects together the recommendations made in the chapter into a comprehensive set.

13.2 National migration statistics

National Migration Statistics are part of the programme of Topic Reports laid before Parliament after each census and published by Her Majesty's Stationery Office. They provide the most detailed cross-classifications of migrants for countries, regions and a set of zones covering the whole of Great Britain. Separate statistics have been produced for Northern Ireland.

13.2.1 Harmonisation and integration of national migration statistics

The 1991 Census National Migration Statistics (OPCS and GROS 1994a, 1994b and 1994c) are organised in two parts occupying three bulky published volumes of nearly 1200 pages of statistics. The tables cover Great Britain only. A separate volume on Migration was published on Northern Ireland (CONI 1994). Harmonisation and integration of these into a volume covering the United Kingdom is needed.

Recommendation M6. *The National Migration Statistics should be consolidated into a* ***United Kingdom product*** *by harmonising and integrating the Great Britain and Northern Ireland tables.*

The content and organisation of the National Migration Statistics (Great Britain) are set out in Tables 13.1 and 13.2. The table titles, the cross-classification used, the geography adopted and the applicable population base for each table are described. This is not the place to debate the utility of each table but to extract some general points. It is useful to make a distinction between *migration area tables* and *migration flow tables*. The National Migration Statistics, the Regional Migration Statistics and the Special Migration Statistics all contain both kinds of table.

13.2.2 Migration area tables

Migration Area Tables classify residents in an area who report migration in the year prior to the census by a variety of attributes. An example is shown in Table 13.3 which shows a small part of National Migration Statistics, Part 2, Table 6. For 32 unit geographical classification proposed by Rees (1989), resident migrants aged 16 and over are counted by sex, distance of move class and socio-economic group. The table can be used to examine which SEGs move most over short distances and which over medium and long distances.

Where Migration Area Tables use the *Type of move (TYMO)* classification they organise migrants by general flow categories within which they are further classified. An example is shown in Table 13.4 which contains an extract from the National Migration Statistics, Part 1, Table 3. The type of move classification varies cleverly with the geographical scale of the focus (called area x below) but is basically of the form

Table 13.1: Organisation and Contents of the National Migration Statistics, 1991 Census, Part 1 (100%)

N	Title	Contents (crosstab variables)	Geography (see Notes)	Migrants (different address one year before census)
1	**Origins and destinations**	*sex by area of origin by area of residence at Census*	Geography One	RESIDENTS
2	**Age and sex**	*sex by area of origin by age by area of residence at Census*	Geography One	RESIDENTS
3	**Marital status by age**	*age by sex by marital status by type of move*	Geography One	RESIDENTS
4	**Single years of age and marital status**	*age (single years) by sex by marital status by type of move*	Countries	RESIDENTS
5	**Economic position**	*sex by economic position by age by type of move*	Countries	RESIDENTS 16+
6	**Migrant households**	*type of move (for migrant and wholly moving households) by economic position of household head, amenities, car availability and persons per room by tenure, not in self contained accommodation and households with a least one child under one*	Countries	HOUSEHOLDS WITH RESIDENTS
7	**Migrants in households**	*type of move by economic position of migrants aged 16 and over, car availability and persons per room for all migrants, by tenure, not in self contained accommodation*	Countries	RESIDENTS IN HOUSEHOLDS
8	**Composition of wholly moving households**	*type of move by household composition*	Countries	WHOLLY MOVING HOUSEHOLDS
9	**Migrants in wholly moving households**	*type of move by age and limiting long-term illness*	Countries	RESIDENTS IN WHOLLY MOVING HOUSEHOLDS
10	**Distance of move**	*sex by distance of move by age*	Geography One	RESIDENTS
11	**Ethnic group**	*sex by age by ethnic group by type of move*	Countries	RESIDENTS
12	**Communal establishments**	*sex by type of establishment by limiting long-term illness by age by type of move*	Countries	RESIDENTS (NON-STAFF) IN COMMUNAL ESTABLISH-MENTS

Source: OPCS and GROS (1994a 1994b)

Notes:
Geography one: Great Britain, England and Wales, England, regions, metropolitan counties, Inner London, Outer London, other main urban centres, regional remainders, Wales, Cardiff District, remainder of Wales, Scotland, main urban centres, remainder of Scotland (32 zones plus aggregations)
Countries: Great Britain, England and Wales, England, Wales, Scotland

Table 13.2: Organisation and Contents of the National Migration Statistics, 1991 Census, Part 2 (10%)

N	Title	Contents (crosstab variables)	Geography (see Notes)	Migrants (different address one year before census)
1	**Origin and SEG**	*sex by area of origin by area of residence by economic position and SEG of employees, self-employed and unemployed*	Countries	RESIDENTS 16+
2	**Age and SEG**	*socio-economic group by sex by age by type of move*	Countries	RESIDENTS 16+ EMPLOYEES, SELF-EMPLOYED AND UNEMPLOYED
3	**Economic position and employment status**	*sex by economic position by employment status by type of move*	Countries	RESIDENTS 16+
4	**Occupation**	*sex by standard occupational classification (sub-major groups) by type of move*	Countries	RESIDENTS 16+ EMPLOYEES, AND SELF-EMPLOYED AND UNEMPLOYED
5	**Industry**	*sex by industry divisions by type of move*	Countries	RESIDENTS 16+ EMPLOYEES AND SELF-EMPLOYED
6	**Distance of move and SEG**	*sex by distance of move by economic position and socio-economic group of employees, self-employed and unemployed*	Geography one	RESIDENTS 16+, ECONOMICALLY ACTIVE
7	**Distance of move and occupation**	*sex by distance of move by standard occupational classification (sub-major groups)*	Geography one	RESIDENTS 16+ EMPLOYEES, AND SELF-EMPLOYED AND UNEMPLOYED
8	**Distance of move and industry**	*sex by distance of move by industry division*	Geography one	RESIDENTS 16+ EMPLOYEES, AND SELF-EMPLOYED AND UNEMPLOYED

Source: OPCS and GROS (1994c)

Notes:
Geography one:
Great Britain, England and Wales, England, regions, metropolitan counties, Inner London, Outer London, other main urban centres, regional remainders, Wales, Cardiff District, remainder of Wales, Scotland, main urban centres, remainder of Scotland (32 zones plus aggregations)

Geography countries:
Great Britain, England and Wales, England, Wales, Scotland

Table 13.3: Portion of Table 6, Distance of move and SEG, National Migration Statistics, Part 2

Residents aged 16 and over with a different address within Great Britain one year before census									
TOTAL MIGRANTS	Economically active			Economically inactive					Sex and distance of move (kilometres)
	Total	On a government scheme	Students	Total	Students	Permanently sick	Retired	Other inactive	
a	b	c	d	e	f	g	h	i	j
				GREAT BRITAIN					
183,597	**154,731**	**2,237**	**912**	**28,866**	**8,710**	**6,549**	**11,743**	**1,864**	**All males**
98,146	82,392	1,257	430	15,754	4,004	4,091	6,827	832	0-4 km
23,768	20,706	270	96	3,062	791	781	1,299	191	5-9 km
16,288	14,293	154	47	1,995	530	511	827	127	10-19 km
12,278	10,445	122	61	1,833	582	405	669	177	20-49 km
5,901	4,776	69	55	1,125	473	169	359	124	50-79 km
27,216	22,119	365	223	5,097	2,330	592	1,762	413	80 or more km
196,134	**120,071**	**1,581**	**1,085**	**75,433**	**8,258**	**5,900**	**20,054**	**41,221**	**All females**
109,192	64,631	919	544	44,561	3,951	3,738	12,204	24,668	0-4 km
25,289	16,553	186	111	8,736	805	752	2,332	4,847	5-9 km
17,043	11,630	128	83	5,413	587	518	1,414	2,894	10-19 km
12,614	8,433	102	92	4,181	529	327	1,105	2,220	20-49 km
5,821	3,667	35	51	2,154	435	121	549	1,049	50-79 km
26,175	15,787	211	204	10,388	1,951	444	2,450	5,543	80 or more km

Source: OPCS and GROS 1994b, p.114. 1991 Census, Crown Copyright.
ESRC/JISC Purchase. Computer readable files provided by the Census Dissemination Unit, Manchester Computing on MIDAS service under */db/census91/topic/nm*.csv.*

Table 13.4: Extracts from Table 3, Marital Status and Age, National Migration Statistics, Part 1 for Yorkshire & Humberside Region

Residents with different address one year before census									
Age	TOTAL MIGRANTS	Males				Females			
		Total	Single	Married	Widowed or divorced	Total	Single	Married	Widowed or divorced
a	b	c	d	e	f	g	h	i	j
Migrants resident in Yorkshire & Humberside Region									
All ages 1 and over	442,196	218,340	122,560	75,491	20,289	223,856	114,355	76,475	33,026
1-4	35,852	18,494	18,494			17,358	17,358		
:									
90 and over	2,101	359	21	91	247	1,742	229	71	1,442
Migrants moving within Yorkshire & Humberside Region									
All ages 1 and over	346,817	168,372	92,120	59,534	16,718	178,445	89,671	60,438	28,336
1-4	28,780	14,782	14,782			13,998	13,998		
:									
90 and over	1,842	312	17	85	210	1,530	200	65	1,265
Migrants moving within districts of Yorkshire & Humberside									
All ages 1 and over	286,949	138,647	76,502	48,395	13,750	148,302	75,110	49,199	23,993
1-4	24,864	12,741	12,741			12,123	12,123		
:									
90 and over	1,505	252	13	65	174	1,253	158	52	1,043
Migrants moving into Yorkshire & Humberside Region from rest of Great Britain									
All ages 1 and over	55,805	28,893	17,602	9,579	1,712	26,912	14,378	9,913	2,621
1-4	3,789	1,950	1,950			1,839	1,839		
:									
90 and over	95	20	1	2	17	75	10	0	65
Migrants moving into Yorkshire & Humberside Region from outside Great Britain									
All ages 1 and over	18,540	9,439	5,156	3,940	343	9,101	4,610	4,029	462
1-4	1,479	790	790			689	689		
:									
90 and over	5	3	1	1	1	2	0	0	2
Migrants moving into Yorkshire & Humberside Region from origin not stated									
All ages 1 and over	21,034	11,636	7,682	2,438	1,516	9,398	5,698	2,095	1,607
1-4	1,804	972	972			832	832		
:									
90 and over	159	24	2	3	19	135	19	6	110
Migrants moving into Yorkshire & Humberside Region to rest of Great Britain									
All ages 1 and over	54,865	27,919	17,598	8,663	1,658	26,946	15,274	9,119	2,553
1-4	3,272	1,670	1,670			1,602	1,602		
:	:	:	:	:	:	:	:	:	:
90 and over	111	16	0	4	12	95	7	5	83

Source: OPCS and GROS 1994a, pp.340-343. 1991 Census, Crown Copyright. ESRC/JISC Purchase. Computer readable files provided by the Census Dissemination Unit, Manchester Computing on MIDAS service under */db/census91/topic/nm*.csv.*

a. Migrants resident in area x
b. Migrants moving within area x
c. Migrants moving within sub-areas of area x
d. Migrants moving into area x from the rest of the country
e. Migrants moving into area x from outside the country
f. Migrants moving into area x from origin not stated
g. Migrants moving from area x to the rest of the country

13.2.3 Migration flow tables

Migration Flow Tables provide interaction information about both origins and destinations. Flow tables can as a result be very large. The number of counts in area tables varies with N, the number of areas, but the number of counts in flow tables varies with N^2. As a result only two of the twelve 100% tables are flow tables (National Migration Statistics Part 1, Tables 1 and 2), and one out of eight 10% tables (National Migration Statistics, Part 2, Table 1). The flow tables spread over many pages and as a result it is difficult to grasp their structure. It would be better to present smaller, more readable tables in the published reports and to place larger tables in spreadsheet files which can be more easily inspected, comprehended and employed in further analysis.

13.2.4 Geographical classifications used in the National Migration Statistics

Two geographical classifications are used in the National Migration Statistics. In 13 tables a simple classification into component countries within Great Britain is used. In 7 tables a 32 zone classification system originally suggested by Rees (1989) is used that identifies large urban centres and region remainders using selected districts and counties. The aim of this classification was to enable researchers to examine the migrant flows into and out of large urban centres. In the event, researchers preferred to use their own classifications built up from ward and district migration matrices for this job (Champion and Dorling 1994; Rees, Durham and Kupiszewski 1996). From the 32 by 32 matrices in National Migration Statistics 1 and 2 researchers can extract interregional matrices and inter-country matrices, but not matrices for inter-county flows or flows between any other comparable classification at the same scale (e.g. NUTS 2 region). The Special Migration Statistics had to be used to generate those flow tables. There is a need to produce a wider set of migration flow tables as part of the National Migration Statistics.

Recommendation M7. *Plan to produce **a much wider set of computer readable flow tables** in the National Migration Statistics using a suite of geographies, ranging from standard regions, through counties (or equivalent) to local government units (unitary authorities, districts).*

13.2.5 The need to publish in computer file form

It would be impossible to provide all the area tables and flow tables that users might wish to use as printed tables. It is doubtful whether repeating the exercise of devoting nearly 300 printed pages to one origin-destination-age-sex array (National Migration Statistics, Part 1, Table 2) is a cost effective way of providing flow data. This was partially recognised in the 1991 Census with the abandonment of the publication of the Regional Migration Statistics as HMSO volumes and their release as computer readable data only. The National Migration Statistics from the 1991 Census were made available later in computer format for manipulation and analysis. The academic community copies are stored on the MIDAS system of Manchester Computing as a set of files under directory */db/census91/topic* with names in the generic form *nm*.csv*, where csv stands for comma separated variables. The *csv* files load easily into spreadsheet packages. Each table, however, spreads over several files and the user has to merge and edit these to obtain a convenient spreadsheet. Alternatively, spreadsheet files are also available from Census Customer Services of the Office for National Statistics.

The nature of the National Migration Statistics needs therefore to be re-thought. The published volume in 2001 should provide definitions, explanations and summary tables but the detailed tables can be published as a library of computer file form on CD-ROM. The number, format and contents of the files need thorough discussion between the census offices and census users.

Recommendation M8. *Plan the National Migration Statistics publication as **a single volume** containing selected summaries, definitions, explanations and analysis and a catalogue of the suite of more detailed tables placed on **an accompanying CD ROM**.*

13.3 Regional migration statistics

Tables 13.5 and 13.6 provide information on the content and geography of the Regional Migration Statistics from the 1991 Census. Table 1 in the Regional Migration Statistics, Part 1 (100%) provides migration flows between and within districts within each region. A typical regional table for Yorkshire and Humberside is shown in Table 13.7. Table 2 in the Regional Migration Statistics, Part 1 provides flows from origins outside the region to districts within the region, using the 32 zone system. Table 3 in the Regional Migration Statistics, Part 1, provides flows to destinations outside the region, using the 32 zone system (geography one), from districts within the region. The other 100% Regional Migration Statistics parallel the National Migration Statistics for those of the 32 zones within the region. The 10% Regional Migration Statistics also parallel the National Migration Statistics using the 32 zone system.

The Regional Migration Statistics from the 1991 Census were originally to be published as a set of volumes, one per region (as from the 1971 and 1981 Censuses) but the decision was made to scrap paper publication because of cost and they were published as computer readable tables. The academic community copies are stored on the MIDAS system of Manchester Computing as a set of files under directory */db/census91/regmig* with names in the generic form *xy*.csv*, where the letters *xy* are replaced by a two letter abbreviation for the region and *csv* stands for comma separated variables. The *csv* files load easily into a spreadsheet package such as MSExcel. Each table, however, spreads over several files and the user has to merge and edit these to obtain a convenient spreadsheet. Table 13.7 shows on one page the inter-district matrix for Yorkshire and Humberside merged to form an easy to read and interpret spreadsheet. Alternatively, spreadsheet files are also available from Census Customer Services of the Office for National Statistics.

Table 13.5: Organisation and Contents of the Regional Migration Statistics, 1991 Census, Part 1 (100%)

N	Title	Contents (crosstab variables)	Geography for area of residence (see Notes)	Migrants (different address one year before census)
1	**Migrants within [region]**	*sex by area of origin [in the region] by area of residence[in the region] at Census*	Regions, counties, districts	RESIDENTS
2	**Origins**	*sex by area of origin [outside region] by area of residence at Census [in the region]*	Regions, counties, districts	RESIDENTS
3	**Destinations**	*sex by area of origin [in the region] by area of residence at Census [outside the region]*	Geography One	RESIDENTS
4	**Age and employment**	*sex by age and employment status by type of move*	Counties, districts	RESIDENTS
5	**Economic position**	*sex by economic position by age by type of move*	Geography one	RESIDENTS 16+
6	**Migrant households**	*type of move (for migrant and wholly moving households) by economic position of household head, amenities, car availability and persons per room by tenure, not in self contained accommodation and households with a least one child under one*	Geography one	HOUSEHOLDS WITH RESIDENTS
7	**Migrants in households**	*type of move by economic position of migrants aged 16 and over, car availability and persons per room for all migrants, by tenure, not in self contained accommodation*	Geography one	RESIDENTS IN HOUSEHOLDS
8	**Composition of wholly moving households**	*type of move by household composition*	Geography one	WHOLLY MOVING HOUSEHOLDS
9	**Migrants in wholly moving households**	*type of move by age and limiting long-term illness*	Geography one	RESIDENTS IN WHOLLY MOVING HOUSEHOLDS
10	**Ethnic group**	*sex by age by ethnic group by type of move*	Geography one	RESIDENTS
11	**Communal establishments**	*sex by type of establishment by limiting long-term illness by age by type of move*	Geography one	RESIDENTS (NON-STAFF) IN COMMUNAL ESTABLISH-MENTS

Source: OPCS and GROS (1991, 1995a)

Notes:
Geography one: Great Britain, England and Wales, England, regions, metropolitan counties, Inner London, Outer London, other main urban centres, regional remainders, Wales, Cardiff District, remainder of Wales, Scotland, main urban centres, remainder of Scotland (32 zones plus aggregations)

Table 13.6: Organisation and Contents of the Regional Migration Statistics, 1991 Census, Part 2 (10%)

N	Title	Contents (crosstab variables)	Geography for area of residence (see Notes)	Migrants (different address one year before census)
1	**Origin and SEG**	*sex by area of origin by area of residence [in the region] by economic position and SEG of employees, self-employed and unemployed*	Geography one	RESIDENTS 16+ EMPLOYEES, SELF-EMPLOYED AND UNEMPLOYED
2	**Destination and SEG**	*sex by area of residence [outside the region] by area of origin [in the region] by economic position and SEG of employees, self-employed and unemployed*	Regions, Wales, Scotland	RESIDENTS 16+ EMPLOYEES, SELF-EMPLOYED AND UNEMPLOYED
3	**Economic position and employment status**	*sex by economic position by employment status by type of move*	Geography one	RESIDENTS 16+
4	**Occupation**	*sex by standard occupational classification (sub-major groups) by type of move*	Geography one	RESIDENTS 16+ EMPLOYEES, AND SELF-EMPLOYED AND UNEMPLOYED
5	**Industry**	*sex by industry divisions by type of move*	Geography one	RESIDENTS 16+ EMPLOYEES AND SELF-EMPLOYED

Source: OPCS and GROS (1991, 1995b)

Notes:
Geography one:
Great Britain, England and Wales, England, regions, metropolitan counties, Inner London, Outer London, other main urban centres, regional remainders, Wales, Cardiff District, remainder of Wales, Scotland, main urban centres, remainder of Scotland (32 zones plus aggregations)

Table 13.7: Regional Migration Table 1, Origins and destinations, for Yorkshire and Humberside, reorganised into one matrix

Source: OPCS and GROS (1995a). 1991 Census Crown Copyright. ESRC/JISC Purchase, supplied by the Census Dissemination Unit, Manchester Computing on the MIDAS service in files */db/census91/regmig/yh1a.csv, yh1b.csv, yh1c.csv, yh1d.csv, yh1e.csv*

Table 1 Migrants within Yorkshire & Humberside Region	Yorkshire & Humberside Region, metropolitan counties, counties, districts																						
Residents of Yorkshire & Humberside Region with different address in Yorkshire & Humberside Region one year before census																							
	Area of residence at census																						
Area of origin and	YORKSH	South Yorkshire Metropolitan County					West Yorkshire Metropolitan County					Humberside County							Humberside County - continued			North Y	
	HUMBERSIDE REGION TOTAL	Total	Barnsley	Doncas	Rother	Sheffiel	Total	Bradfor	Calderd	Kirklees	Leeds	Wakefi	Total	Booth	Cleetho	East Yo	East Yo Borough of Beverley	Glanfor	Great Grimsby	Holderr	Kingsto Upon Hull	Scunth	Total
Males																							
YORKSHIRE & HU	168372	43838	7192	9752	7694	18900	70141	16480	6858	12524	24211	10068	30129	2033	2174	3336	3275	2164	3153	1727	10125	2141	24265
South Yorkshire M	43921	41908	6885	9050	7703	18363	1044	150	60	164	303	347	555	150	41	88	49	35	28	25	91	48	414
Barnsley	6630	6487	6065	128	151	143	368	27	18	60	56	187	60	9	3	21	5	1	5	3	9	4	76
Doncaster	9592	9349	116	8533	217	83	245	31	17	34	78	84	287	119	15	33	16	15	8	7	30	24	131
Rotherham	7644	7590	205	269	6597	523	103	17	5	19	35	27	82	11	4	15	7	11	3	6	17	8	63
Sheffield	19495	16862	309	133	801	17639	328	75	20	51	133	49	146	11	19	19	21	8	12	9	35	12	139
West Yorkshire M	71043	1105	408	309	89	263	87060	15905	6634	12167	22820	9384	781	105	28	272	84	36	25	71	135	25	2097
Bradford	16759	122	17	43	13	49	16044	14469	354	352	760	79	150	9	1	60	19	8	6	18	25	3	443
Calderdale	6813	56	25	11	5	15	6621	244	5867	350	107	33	48	8	2	29	1	2	1	1	4	0	88
Kirklees	12441	191	86	41	15	49	11952	264	277	10785	380	246	150	21	10	37	13	10	7	12	32	6	148
Leeds	25134	344	75	92	40	137	23375	840	129	438	21168	800	301	35	12	99	33	14	7	26	65	10	1114
Wakefield	9896	382	205	122	26	30	9066	58	37	242	505	8226	132	33	3	47	18	2	2	14	9	4	304
Humberside Count	29629	456	42	234	68	113	518	85	33	61	239	92	27987	1614	2096	2603	3049	2060	3090	1597	9802	2056	666
Boothferry	1903	124	5	97	15	7	64	10	3	3	32	16	1596	1340	3	40	56	61	3	6	43	44	117
Cleethorpes	2179	23	4	5	9	5	30	4	2	8	12	4	2121	6	1360	9	11	55	626	10	15	9	15
East Yorkshire	2986	51	3	25	8	15	115	23	5	21	45	21	2550	32	4	2262	86	13	1	57	71	4	282
East Yorkshire Bor	3138	35	3	9	6	17	68	13	3	5	32	15	2940	78	4	117	1906	32	3	56	748	0	90
Glanford	1964	28	5	16	3	5	20	1	0	3	13	3	1890	30	63	11	18	138	36	6	33	300	25
Great Grimsby	3188	30	5	13	8	13	37	10	6	4	10	7	3094	2	819	5	6	43	2385	0	25	9	18
Holderness	1672	14	6	3	2	3	48	6	6	6	17	11	1572	14	7	44	13	10	2	1117	276	0	38
Kingston upon Hul	10290	98	8	37	15	30	108	12	5	16	63	12	10000	54	6	92	804	49	15	342	8561	15	84
Scunthorpe	2287	45	3	29	3	10	28	4	3	3	15	3	2202	51	6	3	5	413	17	3	29	1675	12
North Yorkshire Co	23778	367	47	150	60	110	1519	340	81	104	749	245	825	164	9	373	93	33	10	34	97	12	21066
Craven	1464	10	3	2	0	5	214	153	10	4	43	4	11	0	0	6	2	0	2	0	1	0	1229
Hambleton	2326	39	4	18	6	9	80	15	10	13	33	8	66	17	4	16	15	2	3	0	7	2	2141
Harrogate	4867	60	6	23	10	21	402	73	11	22	265	31	76	7	2	30	11	5	1	7	13	4	4327
Richmondshire	1491	37	1	15	5	16	68	10	14	6	30	6	31	6	0	7	6	1	0	5	4	0	1356
Ryedale	2763	34	4	13	4	13	95	17	11	11	46	10	142	7	0	69	8	6	2	6	11	0	2497
Scarborough	3819	57	11	14	11	15	146	27	5	21	66	27	142	11	2	78	11	8	0	6	25	1	3474
Selby	2798	68	11	42	7	8	354	24	10	17	163	140	227	96	1	73	19	6	2	9	15	6	2149
York	4246	62	7	23	9	23	160	21	10	8	103	18	128	20	0	64	19	2	0	1	21	1	3898

Source: OPCS and GROS (1995a). 1991 Census Crown Copyright. ESRC/JISC Purchase, supplied by the Census Dissemination Unit, Manchester Computing on the MIDAS service in files */db/census91/regmig/yh1a.csv, yh1b.csv, yh1c.csv, yh1d.csv, yh1e.csv*

Recommendation M9. *Plan the **Regional Migration Statistics publication as a single volume** containing selected summaries, definitions, explanations and analysis and a catalogue of the suite of more detailed tables placed on **an accompanying CD ROM**. Alternatively, the National and Regional Migration Statistics could be planned as an integrated volume.*

13.4 Local base statistics and small area statistics

There is relatively little migration information incorporated in the Local Base Statistics (LBS) or Small Area Statistics (SAS). Tables L15, L16 and L17 from the LBS and Tables S15, S16 and S17 from the SAS report on migrants resident in Local or Small areas. Tables L15 and S15 provide an age by sex by type of move classification. Tables L16 and S16 report on wholly moving households by household composition and type of move. Tables L17 and S17 simply report on numbers of migrants among residents in areas by ethnic group. There is considerable scope for extending the crossclassifications of migrants to cover all the dimensions tackled in the National and Regional Migration Statistics, perhaps with broader coding of the variables.

The type of move (TYMO) classification is developed more elaborately at smaller scales. In the Local Base Statistics (Table L15) the following TYMO classification is used

a. Moved within wards
b. Between wards but within district
c. Between districts but within county
d. Between counties but within region
e. Between regions or from Scotland
f. From outside GB
g. Between neighbouring districts
h. Between neighbouring counties/Scottish regions.

The classification is perhaps over elaborate: flows g and h are not needed to complete the picture of in-migration. More importantly "Migrants from the area of residence to the rest of GB" are omitted. There are also no statistics on migrants with origin not stated. These have been merged into the "Moved within wards" category. It is therefore not possible to measure the balance of inflowing and outflowing internal migrants properly. The tables cannot easily be used in population change analysis or population estimation. Out-migrant flows are missing because of the way the LBS and SAS were processed area by area. It is impossible to count out-migrants from an area until the whole national census has been processed. If the "one number census" approach is adopted for the 2001 Census, out-migrants can be counted for all areas, including those of the smallest scale.

Recommendation M10. *Improve the provision of **Migration Area Statistics** by tabulating out-migrant flows, by adopting a simplified type of move classification and by expanding considerably the crossclassifications of migrants. These could be provided as additional tables in the general area statistics produced from the census or as part of the Special Migration Statistics.*

13.5 Special migration statistics

13.5.1 Content and organisation

To meet the need for flow data which users could aggregate to their own requirements the Census Offices in 1981 and in 1991 developed a system of Special Migration Statistics (SMS). Full details of the organisation and content of the SMS are given in OPCS and GROS (1993a, 1993b, 1993c), in Flowerdew and Green (1993) and in Rees and Duke-Williams (1995). These proved to be large and complex data sets subject to considerable suppression. Most users found the statistics very difficult to use, though the 1991 situation was much improved compared with the 1981. In fact, even within the academic community research using these migration data has relied for data extraction on the expertise of a small number of experts at Manchester Computing, and the geography departments at the Universities of Leeds and Newcastle upon Tyne.

In principle, the SMS from the 1991 Census as purchased by ESRC/JISC for academic research, can regarded as a pair of three dimensional arrays. The array dimensions are respectively origins, destinations and attributes, while the cell contents are counts of migrants falling in the origin-destination-attribute combinations.

The *Set 1* array has a very large origin-destination face or matrix. The origins are composed of four types: (1) 9930 wards in England and Wales, (2) 1003 pseudo postcode sectors in Scotland, (3) 96 areas outside Great Britain (individual countries or groupings of countries or other areas such as Northern Ireland), and (4) an origin-not-stated category. The destinations are made up of the first two types of area. The third dimension is made of broad age-sex groups (Table SMS M01) and the categories of wholly moving households and residents in such households (Table SMS M02).

The *Set 2* array has a much smaller origin-destination face or matrix. The origins consist of three types of area: (1) 459 local government districts in Great Britain, (2) 96 areas outside Great Britain (as in Set 1) and (3) an origin-not-stated category. The attribute dimension can be divided into two parts. The first part consists of the broad age-sex groups and wholly moving household categories of Set 1 plus a more detailed age-sex classification (19 age groups by sex). These data are published without modification and are extremely valuable for demographic analysis, population change analysis, population estimation and forecasting.

13.5.2 The confidentiality issue

The second set of attributes by which the Set 2 district migration flows are classified consist of a further 38 to 40 counts organised in 8 or 9 tables (the exact statistics available differ by country). These data were regarded by the Census Offices, despite the very broad coding used, to be a threat to the confidentiality of individual census microdata. As a result, these data were suppressed when the total number of migrants between an origin and destination was fewer than ten. This meant that, although a majority of migrants were reported in these tables, data for only a minority of flows were available. No sensible analysis or aggregation of these flow data could be carried out. Any analysis had to be confined to examining the total flows into and out of districts by these socio-economic characteristics (Champion 1996).

The difficulties posed by this wholesale suppression in the second part of the SMS Set 2 challenged Rees and Duke-Williams (1995b, 1997) to "reverse engineer" the suppression and to reconstruct the flow array virtually in its entirety. This new version of the SMS Set 2 has been made available for general use on the MIDAS service of Manchester Computing with

the *smstab* interface software. The new flow data were used to analyse the pattern of inter-area migration by broad ethnic group (Rees and Duke-Williams 1995c).

So what should be done about the confidentiality problem if equivalent data sets are produced from the 2001 Census. The alternative protection devices which have been suggested include (1) a small amount of record swapping in master database, as part of the one number census approach, (2) sampling, as has been used with other census interaction data set, the Special Workplace Statistics, (3) rounding to a small number base and (4) random perturbation. The first two devices are applied before table production while the third and fourth devices are applied after table production. Because origin-destination matrices for wards/sectors and districts contain a majority of very small flows, methods (2), (3) and (4) will all introduce error. Record swapping is being canvassed widely because although it introduces a small amount of error, the errors should be self-cancelling when aggregation is carried out. However, the research to establish the viability of this method has still to be done. Sampling has the advantage of being used successfully in the past and of making possible tables with more detailed socio-economic coding if the precedent of SWS coding is followed (see Flowerdew and Green 1993 for details).

Recommendation M11. *If it is still considered essential to apply additional protection measures to some migration tables at the smaller geographic scales, **sampling** should be used in preference to suppression, rounding or random perturbation.*

13.5.3 An interface to the Special Migration Statistics

The SMS datasets are large and complex in organisation. ESRC/JISC commissioned two information technology tasks to help users use these data (extract subsets for further analysis).

The first task involved design of software to read, quality check and verify the SMS data sets (Rees and Duke-Williams 1995a). The program, *smstab*, written by Oliver Duke-Williams, was then adapted for general use on the MIDAS service of Manchester Computing (details of the software are provided via the *man smstab* command on MIDAS). The user runs *smstab* by attaching a set of options to the command when issued on-line or accumulates these in a script file. A moderate degree of *Unix* knowledge is needed by the user for success. Several look up tables have been prepared for common aggregations of wards/sectors or districts which the user can select or users can develop their own look up tables. The package is particularly suited to the extraction of migrant flow matrices for further analysis. Duke-Williams (1997) has outlined how the software might be developed into a menu driven, interactive interface for use over the World Wide Web.

The second task involved restructuring of the SMS data sets into system files for use with the fast database package *Quanvert* by Sandra Levine of Quantime Ltd. The *Quanvert* package is normally used with very large household record databases. The package is very good at extracting tables associated with sets of origin-destination pairs specified by users, which can be supplied either interactively or in batch mode. Users, however, have found the procedures for generating origin-destination matrices more difficult to master and prefer to use *smstab* for that kind of job.

Consideration should be given to the provision of a library of migration flow matrices for widely used sets of origins/destinations, the structure of which users can immediately grasp and use. A suite of flow matrices at each scale should be available. These could be linked to the flow matrices produced in the National, Regional and Local Migration Statistics.

Recommendation M12. *Once the structure of the National, Regional, Local and Special Migration Statistics is agreed, a user friendly interface to all data sets should be developed building on experience with the 1991 Census migration data.*

13.5.4 General system for flexible tabulation

In previous discussion between the Census Offices and census users, there has been a strong demand for a flexible tabulation system which users can employ to produce their own designed tables. This desire applies with particular force to migration statistics. The fast and efficient software required to generate the tables already exists and is being acquired by the Census Offices. Such standard database software is much easier to use for generating migration tables than the interface software developed for the SMS data sets. The other ingredients needed for a flexible tabulations system are:

(1) a method for assessing the confidentiality risk of a particular table request;
(2) a system which users can use to design their requests before submission (e.g. the software linked to a dummy database resembling the census database)
and
(3) administrative arrangements between ESRC/JISC and the Census Offices for funding table requests in a cost-effective way.

Recommendation M13. *Consideration should be given to the development a **flexible tabulation service** provided by the Census Offices or a designated agency to provide any additional migration tables needed by users.*

Recommendation M14. *To reduce the cost and speed the delivery of a flexible tabulation service, both the **tabulation/analysis software** and a **dummy census dataset** should be released to the academic community.*

13.6. SUMMARY OF RECOMMENDATIONS

This section of the Chapter collects together the recommendations emerging from the discussion in earlier sections of the paper.

The migration questions

Recommendation M1. *Consideration should be given again to extending the migration question to capture infant migrants, as recommended by the Migration Question Sub-Group. If a question revision is ruled out, migration tables should report the migration status of the mother (or other parent/guardian if no mother is resident in the household) for children under one.*

Recommendation M2. *Consideration should be given again to expanding the relevant questions to capture economic position one year ago, as recommended by the Migration Question Sub-Group.*

The migrant undercount

Recommendation M3. *Consideration should be given to the problem of undercounting specific to the migration statistics, because respondents systematically failed to report a different address one year before the census and because respondents who did report a different address failed to report origin details.*

A household classification of migrants

Recommendation M4. *A careful study should be undertaken of ways it will be possible to classify migrants by household status from the 2001 Census, with a view to recommending the kinds of new tables that might be produced. This study should involve collaboration between the Census Offices and customer sectors.*

A classification of student migrants

Recommendation M5. *A careful study should be undertaken of ways it will be possible to classify migrants by student status from the 2001 Census, with a view to recommending the kinds of new tables that might be produced. This study should involve collaboration between the Census Offices and customer sectors.*

National Migration Statistics

Recommendation M6. *The National Migration Statistics should be consolidated into a United Kingdom product by harmonising and integrating the Great Britain and Northern Ireland tables.*

Recommendation M7. *Plan to produce a much wider set of computer readable flow tables in the National Migration Statistics using a suite of geographies, ranging from standard regions, through counties (or equivalent) to local government units (unitary authorities, districts).*

Recommendation M8. *Plan the National Migration Statistics publication as a single volume containing selected summaries, definitions, explanations and analysis and a catalogue of the suite of more detailed tables placed on an accompanying CD ROM.*

Regional Migration Statistics

Recommendation M9. *Plan the Regional Migration Statistics publication as a single volume containing selected summaries, definitions, explanations and analysis and a catalogue of the suite of more detailed tables placed on an accompanying CD ROM. Alternatively, the National and Regional Migration Statistics could be planned as an integrated volume.*

Local Base and Small Area Statistics

Recommendation M10. *Improve the provision of Migration Area Statistics by tabulating out-migrant flows, by adopting a simplified type of move classification and by expanding considerably the crossclassifications of migrants. These could be provided as additional tables in the general area statistics produced from the census or as part of the Special Migration Statistics.*

Special Migration Statistics

Recommendation M11. *If it is still considered essential to apply additional protection measures to some migration tables at the smaller geographic scales, sampling should be used in preference to suppression, rounding or random perturbation.*

Interface to Migration Statistics

Recommendation M12. *Once the structure of the National, Regional, Local and Special Migration Statistics is agreed, a user friendly interface to all data sets should be developed building on experience with the 1991 Census migration data.*

A flexible tabulation service

Recommendation M13. *Consideration should be given to the development a flexible tabulation service provided by the Census Offices or a designated agency to provide any additional migration tables needed by users.*

Recommendation M14. *To reduce the cost and speed the delivery of a flexible tabulation service, both the tabulation/analysis software and a dummy census dataset should be released to the academic community.*

REFERENCES

Census Office Abbreviations: CONI = Census Office for Northern Ireland; ONS = Office for National Statistics; OPCS = Office of Populations and Censuses; GROS = General Register Office Scotland; HMSO = Her Majesty's Stationery Office

Champion A (1996) Internal migration and ethnicity in Britain. Chapter 4 in ONS and Ratcliffe P (ed.) *Ethnicity in the 1991 Census. Volume 3. Social geography and ethnicity in Britain: geographical spread spatial concentration and internal migration.* HMSO, London.

Champion A and Dorling D (1994) Population change for Britain's functional regions, 1951-91. *Population Trends*, 77, 14-23.

Champion A and Fielding A (edited) (1992) *Migration processes and patterns: Volume I: Research progress and prospects.* Belhaven, London.

CONI (1994) *1991 Census of Northern Ireland. Migration tables.* HMSO, Belfast.

Dale A (1997) What do you want from the 2001 Census? *SARs Newsletter* No.10, pp.1-2.

Duke-Williams O (1997) Interacting with interaction data. Paper presented at the Second ESRC Workshop Planning for the 2001 Census, Muriel Stott Conference Room, John Rylands Library, University of Manchester, 17/18 April 1997.

Flowerdew R and Green A (1993) Migration, transport and workplace statistics from the 1991 Census. In A Dale and C Marsh (eds.) *The 1991 Census user's guide.* HMSO, London. Pp.269-294.

Gould M (1997) What sort of SAR geography do we need? *SARs Newsletter* No.9, pp.4-6.

Jones G (1997) Planning the 2001 census: only four years to go. *Population Trends*, 88,31-35.

ONS, Heady P, Smith S and Avery V (1996) *1991 Census Validation Survey: quality report.* HMSO, London.

ONS, GROS and NISRA (1997) *Census News No.38.*

OPCS and GROS (1991) *1991 census user guide 22. Topic statistics. Regional Migration.* Prospectus. OPCS and GROS, Titchfield.

OPCS and GROS (1993a) *1991 Census user guide 35: Special migration statistics, prospectus.* OPCS, Titchfield.

OPCS and GROS (1993b) *1991 Census user guide 51: Special migration statistics, Cell numbering.* OPCS, Titchfield.

OPCS and GROS (1993c) *1991 Census user guide 53: File specification and instructions for submitting orders on floppy disk, special migration statistics.* OPCS, Titchfield.

OPCS and GROS (1994a) *1991 Census Migration: Great Britain Part 1 (100% tables). Volume 1 of 2.* HMSO, London.

OPCS and GROS (1994b) *1991 Census Migration: Great Britain Part 1 (100% tables). Volume 2 of 2.* HMSO, London.

OPCS and GROS (1994c) *1991 Census Migration: Great Britain Part 2 (10% tables).* HMSO, London.

OPCS and GROS (1995a) *1991 Census Regional Migration: Great Britain Part 1 (100% tables).* Computer readable files. OPCS and GROS, Titchfield.

OPCS and GROS (1995b) *1991 Census Regional Migration: Great Britain Part 2 (10% tables).* Computer readable files. OPCS and GROS, Titchfield.

Rees P (1989) Research policy and review 30. How to add value to migration data from the 1991 Census. *Environment and Planning A*, 21, 1363-1379.

Rees P and Duke-Williams O (1995a) The story of the British special migration statistics. *Scottish Geographical Magazine* 111, 13-26.

Rees P and Duke-Williams O (1995b) Methods for estimating missing data on migrants in the 1991 Census. *Working Paper 95/20,* School of Geography, University of Leeds.

Rees P and Duke-Williams O (1995c) Inter-district migration by ethnic groups. Paper presented at the International Conference on Population Geography, University of Dundee, September, 1995.

Rees P and Duke-Williams O (1997) Methods for estimating missing data on migrants in the 1991 Census. *International Journal of Population Geography,* forthcoming.

Turton I (1997) The 2001 SAR geography: wishes, needs, problems and solutions. *SARs Newsletter* No.9, pp.7-9.

CHAPTER 14
COMMENTS ON THE MIGRATION STATISTICS FROM THE 2001 CENSUS

Tony Champion

14.1 Comments on the recommendations of Chapter 13

Recommendation 1 on extending the migration question to cover the under-ones is useful but results need careful handling in tables so that comparability with 1991 is maintained and so that users understand the difference between migrants born in the interval before the census and migrants alive throughout the interval.

Recommendation 2 is supported.

Recommendation 3 exposes the important problem of undercounting investigation of which needs to be incorporated in the census validation survey so that users can make estimates of the undercount. Alternatively, the problem needs to be tackled through the one number census methodology. Clear identification of imputed migrants will be necessary if these are created.

Recommendations 4 and 5 are supported, and are activities on which academic researchers and the Census Offices need to collaborate.

The recommendations on National, Regional and Local migration statistics in which better migration tables would be incorporated in the area statistics are endorsed.

The recommendation on harmonisation of GB and NI outputs is endorsed. Perhaps this could be coupled with careful attention in the tables to the flows from the Irish Republic and reporting of the Irish Census data on migration from the UK to the Irish Republic. These flows are the largest into and out of the UK to any one country.

The confidentiality issue raised in Recommendation 11 needs careful consideration of the options. If sampling is adopted as the protection device, then a large percentage sample (e.g. 50%) should be used.

14.2 Issues

The population base change makes comparison of 2001 and 1991 Census migration quite difficult. If possible, migration tables from the 2001 Census should be designed in such a way as to show separately the new flows generated by the change.

There is an opportunity to design many more interesting crosstabulations of migration status against other characteristics for incorporation in the area statistics. Migrant flows by income status would be very interesting, for example.

CHAPTER 15
THE POTENTIAL USE OF MOVING UNITS IN BRITISH MIGRATION ANALYSIS

Robin Flowerdew

The importance of migration and residential mobility has long been recognised, both for small-area planning and as phenomena in their own right, and detailed tabulations on migration have been included in the last four British censuses. Information and crosstabulations about migrants are included in several volumes of published Census output, in the Local Base Statistics and Small Area Statistics, and in the Special Migration Statistics. Migration is of course closely related to household structure (and to changes in household structure).

The total number of migrants may be regarded as different from the total number of migration events, in that people tend to move together as families and households. This is reflected in the existence of tables both of households with migrant heads and of wholly moving households. The latter tables, as far as I know, have not been used very much. Wholly moving households are defined as households all of whose members (except children under 1) have moved in the year prior to the Census from the same previous address. This of course excluded any households who have moved as a unit to join anybody else, or who have subsequently been joined by somebody else. They cannot be regarded, anyway, as wholly moving households in a literal sense, because the people who moved may have left behind others who were part of their household a year before.

While it is useful to recognise that groups of people move together, then, the 'wholly moving household' tables do no more than arbitrarily select out units which do not necessarily have much in common. What is often of most interest, socially and demographically, are households which are not wholly moving, where some members have met or joined others, and the techniques used in defining wholly moving households can be used, with little extra effort, to define moving units, which are likely to be of considerably more interest and usefulness.

The suggested definition of a moving unit is the set of members of a household who moved from the same address one year before census date. A wholly moving household would therefore be a moving unit, but so would be a group of people who had moved to join a static individual (perhaps a woman with young children moving in with a parent or friend). A single migrant would be a moving unit, but a household formed by a couple moving in together on marriage would consist of two moving units. It is possible that complicated households might include several moving units, but the vast majority are likely to contain only one or two. At a time of increasing rates of cohabitation and marital separation, the concept of moving unit represents a useful measure intermediate between the individual (who in many cases does not behave autonomously) and the household (which suggests a false stability of composition).

Without the need for additional questions in the Census about marital status or household structure one year ago, this innovation would provide a quantitative basis for research on processes of household formation and growth. Exactly what would depend on how data on moving units were incorporated into census output. They would presumably be identifiable in data sets such as the Household SAR and the LS. As far as information for published tables and area statistics output are concerned, the basic requirements would be:

1) number of moving units in a household (and the presence of a static unit)

2) size of moving units in terms of number of people

3) category of moving unit - categories could include:

a) single adult
b) lone adult with dependent children
c) couple without dependent children
d) couple with dependent children
e) dependent children only
f) pensioner
g) two or more adults, not in couple relationships
h) more complex unit

4) characteristics of moving units or key individuals in moving units

a) age
b) ethnicity
c) marital status
d) economic activity
e) tenure
f) distance of move
g) type of move

5) tables showing the number and types of moving units in flow tables (e.g. Special Migration Statistics).

If data on out-migrants will be available for small areas, it is hoped that tables in categories 1 - 4 above will be available for out-migration as well as in-migration.

This change in how data are processed and published would increase its relevance to a number of important policy questions concerning migration, household formation and household structure, without incurring the disadvantages inherent in adding a new question to the census form.

CHAPTER 16
REACTIONS TO PROPOSALS ON THE MIGRATION STATISTICS AND PLANS FOR CONSULTATION TO 2001

Frank Thomas

This Chapter provides initial reactions to the proposals on Migration Statistics on behalf of the UK Census Offices and information on the timetable for work and consultation proposed by the Census Offices covering Geographical matrices (workplace and migration). However, the views presented here do not constitute official Census Office policy.

16.1 Reactions to proposals on migration statistics

Recommendations on harmonisation of outputs across the UK and of migration statistics across different products are endorsed.

Consistent treatment of the origin not stated migrants should be developed. In principle, they should be clearly identified as a separate flow, leaving the method of allocation to known origins up to users, though advice could be offered.

The recommendations about publishing flow matrices in machine readable data in easy to view and use formats will be given serious consideration.

The hierarchical structure of the SMS is likely to be retained, i.e. less detail on the flows between small areas and more detail on the flows between large areas.

The moving units idea is very interesting and needs to be investigated further.

16.2 Timetable for planning and consultation

The following timetable of work and consultation is planned by the Census Offices.

- A Work Package on origin-destination statistics has been designed. Some of its features were summarised in Chapter 11.
- A paper is being prepared for the Outputs Working Group on interaction statistics.
- A research report is being prepared for the end of 1997 within the Census Offices which will spell out the options for the production of interaction statistics from the 2001 Census.
- Specific proposals for products are planned by next Spring.
- Consultation on those products will then occur in the remainder of 1998.
- The 1999 Census Dress Rehearsal will include the testing of output systems.
- Quality assurance of the products will occur after release from the 1999 Census and users will be invited to take part in this.
- Final specification of the products will be decided in 2000.

CHAPTER 17
DISCUSSION AND RECOMMENDATIONS ON THE MIGRATION STATISTICS

17.1 Discussion

Chris Denham (Office for National Statistics) stressed that the Academic sector's business case for questions should demonstrate the need for migration questions. Only if the business case could be established would a question or question modification be included in the 2001 Census. He also commented that the Census Offices were open to proposals about new data sets that the academic community were prepared to finance, but these needed to be brought forward by Spring 1998 to be considered for inclusion in the 2001 Census White Paper.

Rachel Leeser (London Research Centre) commented that the LRC had not purchased the SMS but had instead found it more cost effective to purchase a special table that provided origin-destination flows matrices for London Boroughs by age, sex and ethnic group, to a coding detail not provided in the SMS Set 2 but possible for London because of the large ethnic minority group populations living in the capital. This had lessons for the academic community: many researchers need such tables and would use an on-line tabulation system, should such as system be feasible and cost effective.

17.2 Recommendations on the migration statistics

On conceptual issues
Recommendation M3 on the migrant undercount, Recommendation M4 on household classification of migrants and Chapter 15's moving unit concept and Recommendation M5 on student migrants were issues to be discussed and worked on by the Sub-group on Interaction Statistics of the Outputs Working Group.

On tables
Work should begin on the design of migrant tables for the national, regional, local and special migration statistics. The revisions should include new flows at the local level (out-migrants to the rest of the country), better distance of move tables and further crosstabs.

On protection measures
Where protection measures were needed, the preference should be for record swapping (at a low level e.g. 1%) or sampling (at a high level e.g. 90%). Suppression should not be used. The migration statistics should identify the extent of imputation so that users could use the original data distinguished from the imputed data, unless this compromised the role of imputation in the package of confidentiality measures.

On questions
Add to the economic position, marital status and tenure questions, questions about these statuses one year ago.

The Editor comments: The Census Offices hold out very little hope for inclusion of any new question that was not in the June 1997 test.

PART 5: SUMMARY

CHAPTER 18
SUMMARY OF RECOMMENDATIONS

This section of the Report will be completed when the Report is revised. It will summarise the recommendations emerging from the proposals and discussion in earlier chapters. Inconsistencies and redundancies will be dealt with. There should also be clear identification of the bodies to which the recommendations are directed. Some of the recommendations are directed to the UK Census Offices, some are directed to ESRC and JISC, some are directed to the ESRC/JISC Census Programme Units and some to census researchers. The recommendations as a whole will framed as a package outlining the policy options for ESRC/JISC with respect to the 2001 Census of Population and should be seen as forming, subject to revision and comment by the ESRC Research Resources Board and the JISC, the position from which negotiations with the UK Census Offices will be carried out over the 1998-2000 period. The aim of those negotiations will be to put in place firm agreements, before the 2001 Census is actually carried out, which can form the basis of planning for dissemination of and research using census outputs from the 2001 round.