Area (2002) 34.1, 82–91

Linking censuses through time: problems and solutions

David Martin,* Danny Dorling† and Richard Mitchell‡

*Department of Geography, University of Southampton, Southampton SO17 1BJ

Email: D.J.Martin@soton.ac.uk

†School of Geography, University of Leeds, Leeds LS2 9JT

Email: D.Dorling@geog.leeds.ac.uk

‡Research Unit in Health, Behaviour and Change, The University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG

Email: Richard.Mitchell@ed.ac.uk

Revised manuscript received 17 September 2001

This paper reviews the difficulties encountered when attempting to study social change by comparing data from successive censuses, and describes a system designed to provide integrated online access to data from the 1971, 1981 and 1991 Censuses in Great Britain at http://census.ac.uk/cdu/lct/.

Key words: *Britain, census, population, change, undercount, correction*

Introduction

One of the main interests of census researchers is the study of social change (Marsh *et al.* 1988; Champion 1995). Indeed, one of the most basic questions that should be answerable using decennial censuses is how a place has changed. However, this simple objective is severely hampered by continual changes in census procedures, which produce differences in geography, variables, environment and access mechanisms between successive censuses. This paper describes the results of research aimed at linking data from the 1971, 1981 and 1991 Censuses for England, Wales and Scotland at geographical levels above wards, and delivering these through a web-based interface to all registered academic users of the census datasets. Comparable data from Northern Ireland are not available to the UK academic community and are thus beyond the scope of the work undertaken here.

Inconsistencies between censuses vary from those that cause minor inconvenience, such as changes in the labelling of census areas, to those causing (apparently) intractable statistical problems, such as radical changes to areal boundaries or questions asked, making impossible the production of comparable population information for the same place at two different census dates. While each of these inconsistencies arises from quite understandable circumstances when considered individually, the combined result is to present an almost impenetrable barrier to the social scientist wishing to examine small-area social change using standard census outputs. Examples of significant barriers to the study of change include 1997 Unitary Authority definitions and the consequent reconfiguration of boundaries surrounding major cities, which hamper attempts to monitor greenfield development; the changing definitions of ethnicity between censuses, which challenge attempts to trace the development of minority communities, and the total absence of software providing access to data from all three of the 1971, 1981 and 1991 Censuses for the purposes of local comparison. These are important challenges that, to date, have remained unaddressed, but which are tackled directly in the work described

ISSN 0004-0894

here. There are also more subtle changes to consider. Even if we have comparable boundaries, the settlement known as York in 1801 is not really comparable with York in 2001. Similarly, even if questions are identical, a car in 1971 was a very different (more valuable) possession than in 2001. The latter issues are acknowledged here, but not considered further in this paper. There were a number of specialized output products from the 1991 Census, such as the Longitudinal Survey (LS) (Dale 1993) and Samples of Anonymised Records (SARs) (Marsh 1993), which are fundamentally structured around individuals and households, but the commentary presented here relates to the most widely used datasets, the Small Area Statistics (SAS) and Local Base Statistics (LBS) (Cole 1993) and also to a variety of interaction and lookup table products.

The work aims to provide statistics that will be ready for comparison with 2001 Census results, and goes some way towards addressing each of the major categories of problem identified in the discussion. The rest of the paper is divided into two major sections, the first of which examines four principal areas of change, and the second of which considers approaches to their solution. We conclude with some more general reflections on the analysis of social change using census data, and on the prospects for analysis after 2001.

The problems

Major differences between censuses may be divided into four categories, namely: geography, variables, environment and access mechanisms. This is broader than the set of obstacles identified by Norris and Mounsey (1983) who identified geography, questions and tabulations – the latter two of which are subsumed within our variables category – but to some extent reflects the growing complexity of data collection and access. Geography refers to the spatial division of the country for census data collection and publication. Variables refer to the questions that are asked by each census, and the definitions, classifications and tabulations through which these are translated into specific topics in the census outputs. Environment concerns the social and political context within which each census is conducted, and access mechanisms concern the organizational and technical procedures in which users must engage in order to access the census results.

Geographies

A requirement of each British census is that it produces population data for the geographical units of the contemporary statutory geography, used for local government and administration. This is therefore a major influence on the geographical units for census organization and publication. Unfortunately, there are many processes that cause change in census geography (Martin and Gascoigne 1994). Statutory geography is subject to continual revision (Johnston *et al.* 2001), and this forces changes to the areas for which census data must be reported. The geographical division of England and Wales into enumeration districts (EDs) is concerned primarily with the management of enumerator workloads, and must therefore accommodate residential development, demolition and variations in density and residential structure. Until 2001 in England and Wales these EDs also formed the lowest level of the census output geography, the same areal units being used for both purposes. Only 32 per cent of ED boundaries remained unchanged between 1981 and 1991, making direct analysis of social change at the smallest geographical scales highly problematic.

Martin (1998) outlines new procedures that will see a complete separation of the geographical units for 2001 enumeration and output in England, Wales and Northern Ireland, similar to that already in place in Scotland in 1991. A new set of output areas (OAs) is to be computed once the 2001 Census data have been collected with the aim of achieving maximum consistency with postal geography, while maintaining a degree of control over OA populations, shape and social composition. While having many advantages over the traditional EDs, the new system still does not provide compatibility of census boundaries with those used in previous censuses.

Variables

Social change leads to change in research and policy interests, and there is always demand for a far broader range of questions than can be practically included in the census questionnaire. A full review of changing census questions is provided by Dale (2000). For example, the 1991 Census contained four completely new questions concerning ethnic group, limiting long-term illness, term-time address of students and weekly hours worked. There were also extensions to other questions, for example concerning dwelling type and amenities (Champion 1995). There was much pressure for the 2001 Census to include a question on income, a proposal that

was eventually dropped (Dorling 1999). There are changes to the recording of household composition and employment and additional questions on religion, illness, size of employers' organization, time since last employment and unpaid care (Moss 1999; Dixie and Dorling 2002). There is always a tension between the needs to include new topics that have grown in social importance and to maintain comparability between censuses. It is not possible to examine change until the same question has been asked for two successive censuses. This dilemma is reflected, for example, in the changes that have occurred in the wording of household amenity questions, with access to an inside WC being a most revealing social indicator in 1971 whose utility had declined almost to zero by 1991, when the presence of central heating was added in an attempt to find a more contemporary indicator of basic housing quality.

Once census data have been collected, the answers are coded and processed in a number of complex ways in order to produce the final outputs. Changes to these processes also have significant impacts on the comparability of outputs. With growth in the number of census users and in generally available computing power, the SAS have been produced for more cross-classified cells in each successive census. The 1971 SAS comprised around 1600 cells arranged across 28 tables; by 1981 this had grown to 5500 in 43 tables and in 1991 to 9000 cells in 95 tables (Rhind 1983; Openshaw 1995). Differences in detailed definitions make the task of identifying comparable cells within these tables challenging, even for the experienced census user, and the ten-year intercensal period ensures that in most organizations there are few users with experience of more than two censuses. A further methodological difference when comparing British censuses is that between variables coded at 100 per cent and 10 per cent. Historically, certain questions that are time-consuming to encode, such as those relating to employment, have only been processed for 10 per cent of the census returns, leading to the need for additional consideration of sampling error in the interpretation of these results. For 2001, all variables will be coded on 100 per cent basis for relevant records.

Due to the inevitability of some degree of undercount (discussed in more detail below), a major innovation for 2001 will be the introduction of 'One Number Census' processing, in which a very large Census Coverage Survey will be used to estimate the numbers of individuals and households missed by the census in different neighbourhood types. This information will be used to impute missing individuals and households before the tabulation of the published counts, so that all the census results will relate to 'one number' representing the total enumerated and imputed population of each area (Diamond *et al.* 2002).

Environment

The specific details of any particular census are further affected by a wide range of factors that can be summarized as the 'environment' in which the census is taken, and which can have major impacts on the published data. These add to the complexity of analyzing social change. An example is the decision not to include an income question in 2001. This is attributed to the risk of a negative reaction from respondents damaging the entire data collection exercise, although Dorling (1999) argues that the real reason for not asking this question was a fear by government of revealing social inequality. In the United States, by contrast, the inclusion of an income question is not considered controversial. Similarly, confidentiality constraints that are applied to prevent the inadvertent identification of individuals in the census outputs may be considered as technical responses to external environmental factors: there is no absolute value associated with the protection of individual data. The constraints that are considered necessary here are far in excess of those that pertain in Scandinavia, for example. These general societal and political variations are observable not only internationally, but also between successive British censuses.

Most significant of these factors in 1991 was the missing of about 1.2 million people – commonly attributed to a widespread suspicion that census returns would be used to identify individuals who were seeking to avoid registration for the unpopular community charge in place at the time (Simpson and Dorling 1994). There was also little perceived incentive for people to fill in the 1991 form, given a government since 1979 that had showed little interest in the welfare of many groups. One example of this disinterest in real counting was the 30 revisions made to the official unemployment count between 1979 and 1991. That the political make up of government influences social statistics is made clear in America by Republican disinterest in a correct census. Political influences thus have methodological implications. The analysis of change involving 1991

counts is made enormously complex by the impossibility of knowing what proportion of observed change is due to people missed in 1991 and what proportion reflects genuine social change. This is of particular significance with regard to the more marginalized population sub-groups that were most strongly affected by the undercount, and that we might hope have experienced most change since 1991. Researchers soon to be receiving the 2001 data will be faced with some difficult obstacles when comparing the new (one number census) results with those obtained in 1991.

Further examples of such external factors affecting the 2001 Census include the foot and mouth epidemic and the timing of the general election. The epidemic made impossible the conventional enumeration of many thousands of agricultural addresses, and at one stage threatened the timing of the census itself, although eventually leading to the rescheduling of the general election and thus ironically avoiding a clash with the census.

Access mechanisms

Organizational and technical changes in the intercensal periods lead to very different access mechanisms for census outputs. While a number of paper-based topic and county reports have continued to be prepared for each census, the primary means of dissemination has become electronic. Paper abstracts and tables help the non-expert user but are rather limited in the opportunities they afford for research. The researcher seeking to analyze employment and demography must become conversant with the organization of the data and convert their requirements into area codes, tables and cell numbers. Despite huge intercensal developments in computing, access mechanisms continue to present significant barriers to use. Perhaps the most important obstacle for analysis of change is that the access mechanisms remain for the most part different for each census – both in terms of data registration and/or purchase arrangements, and the software tools required. Rees (1995) describes in detail the arrangements for academic access to data following the 1991 Census, noting that different registration procedures apply to 1981 and 1991 data.

SASPAC (Davies 1995) has been the mainstay of most census users during the 1980s and 1990s, providing initially command-based and then menu-based PC extraction and manipulation of SAS and LBS across all sectors of census users. More recently, CASWEB (Martin *et al.* 1998; Harris *et al.* 2002) has seen considerable use in the academic community, offering web-based delivery of census data, integrated with relevant metadata and mapping. At present CASWEB does not incorporate 1981 data for which SASPAC must still be used. A copy of the 1971 data is held at the ESRC Data Archive in its original punched card format, although this is not easily readable by any standard data extraction software.

Approaches to a solution

In this section, we describe the approaches to each of these problems that have been implemented as part of the Linking Censuses Through Time project. Problems of change are embedded in the census-taking process, and it is not possible to overcome them completely. Nevertheless, this work is grounded on the observation that most social scientists with an interest in using census data tend to think in terms of topics, questions and concepts. They may be interested in geographical systems that are not neat aggregations of the published census geography, and are certainly concerned with measuring change over time, one of the major objectives of a decennial census. Such users wish to access the data infrequently and quickly, and do not wish to acquire familiarity with cell and table structures, changing population bases and geographies. These users therefore need an interface to the census data that incorporates best practice in dealing with each of these issues.

Linking geographies

There are various approaches to the linkage of data for incompatible areal units. These either remodel the data to some underlying surface-based representation (e.g. Bracken and Martin 1995); use areal interpolation to transfer data from one set to another (Goodchild *et al.* 1993), or use lookup tables to make best-fit assignments of one set of areal units into another. Where researchers are concerned with the analysis of change over very small geographical areas, the former approaches may often be the only alternative, but for larger areas, the latter have considerable utility. For the purposes of analyzing intercensal change, higher-level geographical units such as local authority districts, parliamentary constituencies, health authorities and travel to work areas often form the most appropriate units for analysis, and we focus on these units here.

Given the degree of error in 1991 Census data, sophisticated schemes for deriving higher level

Table 1 Subject areas and key variables

Subject	Variables
Demography	Total persons; total males; total females; 11 age categories
Migration, country of birth and ethnicity	Total persons; total migrants; 4 country of birth categories; 4 ethnicity categories
Economic activity	Total persons aged 16 and over; 6 economic activity categories
Tenure and amenities	Total households; 4 tenure categories; overcrowding, lacking basic amenities
Living arrangements and illness	Total persons; in couples; single widowed or divorced, limiting long term illness by 7 age groups
Household composition	Total households; 6 household composition categories
Communal establishments	Residents in communal establishments; 5 communal establishment categories
Industry	Total persons in employment (10%); 5 industry categories
Occupation (SEGs)	Total persons economically active; 17 socioeconomic groups
Travel and transport	Total persons in employment (10%); 7 travel to work categories; total households; 4 car ownership categories
Qualifications	Total persons aged 18 and over (10%); having degree or equivalent

geographies introduce a spurious aura of accuracy. Firstly, approximately 100 people per ward were missed and secondly, geocoding of the 1991 data was poor. Thousands of ED centroids were corrected following their release in 1991 (Dorling and Atkins 1995). Following this correction, 1991 ED data were aggregated to 1981 wards using traditional point-in-polygon techniques and a similar process was applied to 1971 data. Subsequently, these techniques have been used to aggregate all the 1971–1991 data to higher level geographies, such as unitary authorities, parliamentary constituencies and health geographies. For any geographies that comprise less than 1000 areas in Britain, this technique is usually adequate. Errors introduced by these simple aggregation techniques are small when compared to the errors that stem from under-enumeration and those that undoubtedly still remain in the georeferencing of the 1991 data.

The 1981 ward geography is used as the base for all the areal aggregations produced here, and the full 1981 SAS for 10 444 wards represents the most geographically detailed output available. Over 30 different geographical systems have been constructed, ranging in resolution from the 1981 wards to counts for the whole of Britain, and in time from the Registration Districts of 1898 to 2001-based geographies including travel to work areas (TTWAs) and Government Office Regions.

Linking variables

Given the extremely large number of variables that are available from the 1971, 1981 and 1991 SAS, the work described here has focused on the identification of approximately 100 key variables, most of which may be defined in comparable ways between the three censuses. These variables are those that we believe will be comparable with counts from the 2001 key statistics. This has been undertaken by careful interpretation of the SAS definition tables from each of the censuses, as reproduced in Rhind (1983) and Openshaw (1995). The selection of key variables is listed in Table 1, and has involved the addition and subtraction of counts in each of the census SAS tables in order to achieve comparable counts. Each of these censuses used a different population base, and it has not been possible to directly reconcile these. The counts for each year are therefore defined relative to their own population base, and appropriate denominator variables are provided for each subject area. However, see Dorling (1995, Table 1.13) for a justification of why changing population bases may be ignored.

Correcting for the undercount

The single greatest component of this work has been the development and implementation of a procedure for correction of the 1991 counts. We argue that this is the single most important impact of social environment on the 1991 data, and that it is amenable to correction to a large degree. It is beyond the scope of this paper to cover the correction procedure comprehensively, and the reader seeking a more detailed technical presentation is referred to Mitchell *et al.* (2002). The outputs described here are available to registered users of

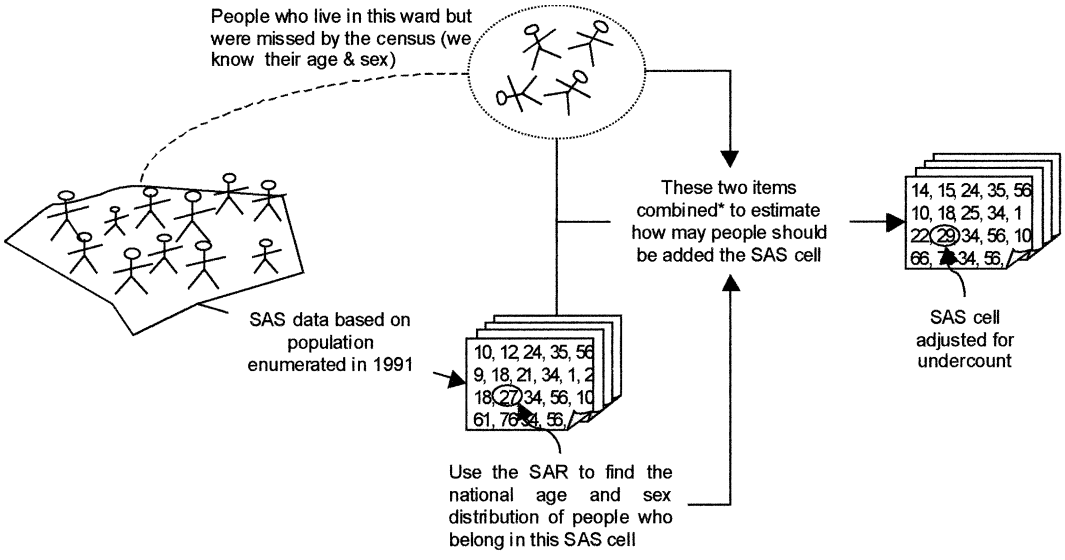

Figure 1 Overview of the 1991 SAS cell correction process

the census data at http://census.ac.uk/cdu/lct/. There will, of course, have been an element of undercounting in 1981 and previous censuses, but this was small in comparison to the 1991 problem, and insufficient ancillary information exists from 1981 to apply detailed correction methods.

In the years following 1991, there has been much discussion among census researchers as to the best way of dealing with the undercount. It is generally agreed that the best way to characterize the distribution of absence from the census is to consider specific age, sex and location groups. A statement from the Census Offices 'that the effects of undercoverage in 1991 are likely to be unimportant for most purposes' (OPCS 1995, 2) has led to widespread use of the SAS without much further consideration of the likely implications. Unfortunately, in research situations in which an accurate count of individuals is essential, this is problematic, and particularly so when the researcher is concerned with change. Such work includes, for example, the calculation of standardized mortality rates for which accurate counts of the denominator population by age/sex groups are required if the resulting death rates are not to be artificially inflated (mortality rates tend to be highest in areas of greatest underenumeration). In response to the undercount, the Census Offices released estimates of the numbers of people missed, and the 'Estimating With Confidence' (EWC) project (Simpson *et al.*, 1997) used this information as the basis for a set of corrected counts by age and sex at the ED/OA level. Extension of the undercount information to include further characteristics of those missed is more difficult, but belief in its importance has led to its forming a major focus of the work described here.

Our correction of 1991 counts rests on the construction of age and sex distributions for each resident-based tabulated cell in the SAS, so that EWC data could be used to estimate the extent to which each cell would be affected by the undercount in each ward, as illustrated in Figure 1. For example, a SAS cell typically containing a young male population will need greater correction than one containing middle-aged females, due to their different proportional presence in the undercounted population. Three sources of information have been used in the implementation of this procedure. The age/sex distribution of each SAS cell has been estimated using the disaggregate sample of individual census returns in the SARs. The EWC data provide estimates of the age/sex counts of persons missed at the ward level, and the SAS provide the base figures to be corrected.

The first stage of the correction process has been the creation of a 40-category age/sex distribution for each SAS cell count, at the national level, by tabulation of the SAR records. A custom-written program called SASGEN was written to achieve this task. This process was time-consuming, but not inherently

Table 2 Adjustments to 1991 Census population counts

Region	1991 census count	1991 census count after correction	Number of people added	Percentage change in population count
North	3 025 486	3 091 104	65 618	2.17
North West	6 240 951	6 394 868	153 917	2.47
Yorkshire and Humberside	4 833 745	4 981 012	147 267	3.05
East Midlands	3 951 486	4 034 798	83 312	2.11
West Midlands	5 147 380	5 264 387	117 007	2.27
East Anglia	2 026 080	2 081 996	55 916	2.76
South East	17 192 728	17 642 550	449 822	2.62
South West	4 605 218	4 716 976	111 758	2.43
Wales	2 832 963	2 890 642	57 679	2.04
Scotland	4 999 303	5 104 090	104 787	2.10
Britain*	54 855 340	56 202 423	1 347 083	2.46

*Total excludes 421 people not assigned to a region by the census package's aggregation system

Table 3 Percentage adjustments to population counts by age group in selected Local Authorities

	Total	15–19	20–24	25–29	30–44	45–49	60–64	65–74	75+
Manchester	8	15	29	16	4	1	0	0	3
Leicester	5	8	23	9	3	0	−1	0	2
Taunton Deane	2	9	−4	3	1	1	0	0	3
Carlisle	1	1	-2	4	1	1	−1	1	2
Gillingham	1	−2	−1	2	1	1	0	0	2
Woking	0	−7	−4	3	2	0	1	0	3

complex: the greatest obstacle lay in the correct interpretation of the SAS cell definitions in order to ensure that SAR individuals were correctly allocated to SAS cells. SASGEN was run once as a validation check, and the resulting counts compared with SAS totals for the whole country, of which they should represent approximately 1 per cent for the 100 per cent tables and 10 per cent for the 10 per cent tables. Following validation and code correction, SASGEN was run again in its entirety to produce a set of values for the next stage of correction.

The second stage has been the combination of the EWC population counts with the existing SAS data, using the SASGEN results, allowing the entire dataset to be 'corrected' to form a new 1991 dataset. An age/sex breakdown was computed for each ward, and numbers missed estimated for each age/sex category. Using the SASGEN national results and the local ward data, the likely age/sex breakdown of each SAS cell in that ward was calculated, and adjustment factors calculated for each category. These were then recombined to determine the overall level of adjustment (if any) required for that cell in that ward. This process has been undertaken for every ward in the country for every 1991 SAS cell.

The net result of this process can be seen in a comparison between the population counts from the 1991 Census and the data 'corrected' for undercount. Table 2 presents this information using Registrar General's standard regions to indicate between-area differences in the degree of undercount correction. The impact of correction can be more clearly seen at local level, when we consider different age groups. In Table 3 we present the change in population counts for various age groups in six local authorities. The authorities have been selected to reflect the typical range of adjustment values. Notice that for some age groups the population count has actually been reduced by the correction process. This reflects discrepancies between the 1991 Census counts and the subsequent EWC population estimates. Reduction in

Table 4 Students living in parliamentary constituencies and not counted in the 1991 Census: top ten by absolute numbers missed

Constituency	1991 census count	Corrected count	Number missing	Percentage missing
Oxford West and Abingdon	5011	15 091	10 080	67
Cambridge	4738	12 608	7870	62
Leeds North West	4460	7924	3464	44
Bristol West	5718	8593	2875	33
Newcastle upon Tyne Central	4925	7654	2729	36
Liverpool Riverside	4118	6407	2289	36
Oxford East	4105	6380	2275	36
Edinburgh South	3566	5829	2263	39
Manchester Central	4236	6395	2159	34
Sheffield Hallam	4255	6372	2117	33
Total for top 10	45 132	83 253	38 121	46

population amongst the younger age groups is likely to reflect the fact that the corrected 1991 counts locate students at their term-time address. This is further illustrated by Table 4, which shows the numbers of students counted by the 1991 Census compared to the corrected counts for the ten Parliamentary Constituencies missing the largest numbers of students. The 2001 Census will count students at their term-time addresses and therefore any attempts to measure change relative to uncorrected 1991 counts will be severely flawed.

By tending to miss people of a certain age and sex, the 1991 Census not only distorted the demographic picture of Britain, it distorted the social and economic picture too. This is of vital significance for social scientists interested in the distribution of key social and economic characteristics (both in social and spatial terms).

Access mechanisms

The system described in this paper is available to registered users of the census data at http://census.ac.uk/cdu/lct/. The interface is developed from a program for accessing census data that was created by Dorling in 1989 and has been used by a small number of other researchers during the 1990s. This is a command-line driven C program that uses 1981 wards as a base geography, and reads external lookup tables to produce run-time aggregations to the requested output geography. The existing program has remained largely unchanged, but additional geographical aggregations have been added to the library and the entire corrected 1991 dataset has been created and added, giving access to 27 057 separate cells in total, although some of these contain identical counts. The base data thus comprise the full 1971, 1981, 1991 (uncorrected) and 1991 (corrected) SAS, held for 1981 wards, with lookup tables to a range of other geographies. An entirely new web interface to the program has been created, adopting a similar approach to that taken for the SURPOP interface to census surface data (Martin *et al.* 1998).

The new interface allows users to specify one or more censuses from 1971 to 1991 (corrected, uncorrected or both). One output geography may be selected for a single program run, and the available geographies are organized by function, and listed in Table 5. The key census variables listed in Table 1 may be selected in any combination, organized by the categories shown in the first column of the table. The organization of data by topic or 'question', rather than by table and cell numbers has been a primary consideration in this work, which we hope will help to widen the census user base among social scientists.

The interface incorporates a database of geography and variable codes appropriate to each census and handles all the required additions and subtractions of the individual cell counts in order to achieve comparable definitions. Metadata are provided in the form of notes on each geography and variable, and also by the provision of the full SAS table layouts for each census. Users may elect to include area names and numbers with their output. The interface builds a command line that may be edited or entered directly by expert users. Further manipulation of the

Table 5 Aggregation geographies

Geography type	Geographical units
Countries and regions	Whole of GB; 1981 countries and regions; 2001 regions
Historical (pre 1951)	Registration districts 1989 and areas 1939
Administrative and aggregations	Local authorities, districts, counties and official aggregations, 1951; 1971; 1981; 1996; 1998; 2001
Electoral	European and parliamentary constituencies, 1993; 1997
Functional	Travel to work areas and derivatives 1981, TTWAs 2001
Health and education	FPC areas; health authorities 2001; LEAs

selected variables is available through a 'calculator', again permitting the expert user to adapt the standard variable definitions or to collapse categories. Once a particular data extraction run has been specified, the user's MIMAS identification is validated in order to confirm their eligibility to access the requested data, and the results are delivered as a comma-separated file for download to the user's web browser.

Conclusions

This paper has considered four major obstacles to analysis of intercensal social change. It has presented the results of research that delivers streamlined access to censuses from 1971 to 1991 incorporating a range of areal units; careful selection of comparable variables; correction of 1991 counts, and web-based delivery of data from a single entry point. The system offers thematic choices to the non-specialist user, based around broad topics common to many social research questions, while still allowing the expert to define precisely the variables that they want without the need for separate access systems for each census. We hope it will become the standard interface through which non-expert (and expert) census users will access data from 1971 through to 2001 for higher level geographies. We have not attempted to incorporate 1961 and 1966 into our current work. These are potential areas for future work, as we begin to use digital censuses as a historical resource.

The new output geography for 2001 will offer a more flexible geographical base than that previously available, comprising more, smaller geographical units that are more easily integrated with postal geography and with other non-census datasets. 2001 OAs will also be used in the publication of a range of new neighbourhood statistics in the intercensal period (Social Exclusion Unit 2000). The address-level geographical databases created by the Census Offices offer the potential for updating and reaggregation of census counts to new statutory areas following the initial publication of results. No degree of advanced planning can guarantee a census process that is immune from the idiosyncracies of the social and political environment, as the foot and mouth epidemic demonstrated. International experience, such as that of the 2000 Census in the United States, would suggest that compliance with government censuses and surveys is declining, but in 2001 in Britain this will be offset by the entirely new ONC approach being taken to validation and imputation of missing individuals and households. In many ways, the most significant development for the future may be the new Census Access developments project (Denham 2000), which will offer far wider and freer access to the results of the census, and will therefore hopefully ensure that there are many more users in 2001 than in any previous decade.

We believe that this work will mean that these new census data users will also now find it easier to compare the 2001 Census with the last three. This census alone will only tell us about the general human geography of the country approximately two years after it is enumerated. We suspect that the general picture will differ only slightly from that produced in 1991, with the centres of cities still tending to be poor, ethnic minority groups still concentrated in London and other urban centres, and unemployment levels down, but with a broadly similar geography. What we do not know and what is arguably most interesting, is how the geography of British society has changed in more subtle ways. In which cities and areas has there been gentrification and residualization? Are certain social groups

migrating away from suburbs? Is Britain becoming more or less segregated in terms of ethnic diversity? These are questions that require a longer-term view and, thus, easier access to the data we have from 1971 onwards. We hope that the work we have presented here will provide the tools to answer these and many more substantive research questions.

Acknowledgements

The research described in this paper has been supported by ESRC award number H507255151. The authors gratefully acknowledge the expert assistance of Jason Sadler in programming the LCT web interface. We also wish to acknowledge Helen Durham for help with constructing the 2001 geographies and Paul Wilkinson for work on making the original census program more portable. Richard Mitchell is now funded by the Chief Scientist Office of The Scottish Executive Health Department (SEHD) and the Health Education Board for Scotland (HEBS). The opinions expressed in this paper are those of the authors not of SEHD or HEBS.

References

Bracken I and Martin D 1995 Linkage of the 1981 and 1991 Censuses using surface modelling concepts *Environment and Planning A* 27 379–90

Champion A G 1995 Analysis of change through time in **Openshaw S** ed *Census users' handbook* GeoInformation International, Cambridge 307–35

Cole K 1993 The 1991 local base and small area statistics in **Dale A and Marsh C** eds *The 1991 Census user's guide* HMSO, London 201–47

Dale A 1993 The OPCS longitudinal survey in **Dale A and Marsh C** eds *The 1991 Census user's guide* HMSO, London 312–29

— 2000 Developments in census taking in the last 25 years *Population Trends* 100 40–6

Davies H 1995 Accessing the data with SASPAC91 in **Openshaw S** ed *Census users' handbook* GeoInformation International, Cambridge 83–109

Denham C 2000 Delivering the 2001 UK Census results: Innovation in dissemination (www.statistics.gov.uk/Census2001/pdfs/chrisdenham.pdf) Accessed 24 May

Diamond I, Cruddas M and Woolford J 2002 The one number census in **Rees P, Martin D and Williamson P** eds *The census data system* Wiley, Chichester 283–92

Dixie J and Dorling D 2002 New questions for the 2001 Census in **Rees P, Martin D and Williamson P** eds *The census data system* Wiley, Chichester 271–82

Dorling D 1995 *A new social atlas of Britain* Wiley, Chichester

— 1999 Who's afraid of income inequality? *Environment and Planning A* 31 571–4

Dorling D and Atkins D J 1995 *Population density, change and concentration in Great Britain 1971, 1981 and 1991* Studies on medical and population subjects 58 HMSO, London

Goodchild M F, Anselin L and Deichmann U 1993 A framework for the areal interpolation of socioeconomic data *Environment and Planning A* 25 383–97

Harris J, Hayes J and Cole K 2002 Web-based access to area statistics in **Rees P, Martin D and Williamson P** eds *The census data system* Wiley, Chichester 107–16

Johnston R, Pattie C, Rossiter D and Dorling D 2001 *From votes to seats: British general elections since 1950* Manchester University Press, Manchester

Marsh C 1993 The sample of anonymised records in **Dale A and Marsh C** eds *The 1991 Census user's guide* HMSO, London 295–311

Marsh C, Arber S, Wrigley N, Rhind D and Bulmer M 1988 Research policy and review 23. The view of academic social scientists on the 1991 UK Census of Population: a report of the Economic and Social Research Council Working Group *Environment and Planning A* 20 851–9

Martin D 1998 Optimizing census geography: the separation of collection and output geographies *International Journal of Geographical Information Science* 12 673–85

Martin D and Gascoigne R 1994 Change and change again: geographical implications for intercensal analysis *Area* 26 133–41

Martin D, Harris J, Sadler J and Tate N 1998 Putting the census on the web: lessons from two case studies *Area* 30 311–20

Mitchell R, Dorling D, Simpson S and Martin D 2002 Bringing the missing million home *Environment and Planning A* 34 in press

Moss C 1999 Selection of topics and questions for the 2001 Census *Population Trends* 97 28–36

Norris P and Mounsey H M 1983 Analysing change through time in **Rhind D** ed *A Census user's handbook* Methuen, London 265–86

OPCS 1995 Census coverage *Census Newsletter* 32 2

Openshaw S ed 1995 *Census users' handbook* GeoInformation International, Cambridge

Rees P 1995 Putting the census on the researcher's desk in **Openshaw S** ed *Census users' handbook* GeoInformation International, Cambridge 27–81

Rhind D ed 1983 *A Census user's handbook* Methuen, London

Simpson S and Dorling D 1994 Those missing millions: implications for social statistics of non-response to the 1991 census *Journal of Social Policy* 23 543–67

Simpson S, Cossey R and Diamond I 1997 1991 population estimates for areas smaller than districts *Population Trends* 90 31–9

Social Exclusion Unit 2000 *Better information* Policy Action Team Report 18 The Stationery Office, London