Sheffield Hallam University

A Sheffield Hallam University thesis

Fines are charged at 50p per ho

REFERENCE

Improving urban and regional policy-making using microdata from UK population censuses.

Christopher Gardiner

Published works submitted in partial

fulfilment of the requirements of

Sheffield Hallam University for the

degree of Doctor of Philosophy

on the basis of published work.

November 2004

Sections

Acknowledgements and Dedications

During the process of completing this submission I have received help, advice and encouragement from many people.

In particular I would like to thank Rob Furbey for his advice and guidance throughout the process, and Russ Haywood for his initial encouragement to undertake the submission. I would like to express my gratitude to Rosalie Hill, my co-author for two of the earliest papers, for her help in developing this area of research. My thanks also to other colleagues at Sheffield Hallam University - in the Schools of Urban and Regional Studies and Environment and Development, and the Centre for Regional Economic and Social Research - and at the Cathie Marsh Centre at the University of Manchester.
The work and effort involved in the research programme could only have been sustained with the support and backing of my family. I would like, therefore, to dedicate this work to my wife, Carole, and to my children, Elizabeth and David.
Finally I hope that this research in the longer term will have made a small contribution to the quality of policy-making and use of resources in the public services.

Christopher Gardiner

November 2004

1) Numbered list of Published Works submitted

1   Gardiner, C, and Hill, R, 1996, 'Analysis of access to cars from the 1991 UK Census Samples of Anonymised Records: a Case Study of the Elderly Population of Sheffield.', *Urban Studies*,33, pp 269-281

2   Gardiner, C, and Hill, R, 1997 'Cycling on the Journey to Work: Analysis of Socio-Economic Variables from the UK 1991 Population Census Samples of Anonymised Records.', *Planning Practice and Research*,12, pp 251-261.

3   Gardiner, C, 1996 'Applications of the SARs in Local Government Policy Making' *British Urban and Regional Information Systems Association,*126, pp 10-12.

4   Gardiner, C, 1996 'Developing Indicators to Assess the Potential for Urban Regeneration - A Case Study of Sheffield Using the 1991 Census of Population Sample of Anonymised Records', *Working Paper No 34, Centre for Regional Economic and Social Research*, Sheffield Hallam University.

5   Gardiner, C, 1998 'Developing Indicators to Assess the Potential for Urban Regeneration: Improvements using the 1991 Census of Population Samples of Anonymised Records', *Urban Studies,* 35, pp1519-1540.

6   Gardiner, C, 1998 'Improvements, by using a new set of data, to the English Standard Spending Assessments: the 1991 Census of Population samples of anonymised records.', *Environment and Planning Series C: Government and Policy,*16,pp 543-558.

7   Gardiner, C, 2002 'Improving information for Regional Development Agencies' strategies and policies: using microdata from the United Kingdom's 2001 Census of Population', *Environment and Planning C: Government and Policy,*20, 39-55

8   Gardiner, C, 2001 'Informing Policy Making: New Approaches to Analysing the 2001 Census', *Local Government Studies,*27, No 4, pp71-88.

NOTE: A further paper has been accepted for publication in Local Government Studies, but was not admissible for this submission under Sheffield Hallam University's regulations. It is, however, of

relevance in the context of later developments in the research programme and is therefore listed below rather than in the general reference section at the end of this submission:

9) Brown, M, and Gardiner, C, 2004 'Informing Policy Making at sub-local authority spatial levels: using Small Area Microdata from the 2001 Census' *Local Government Studies,*30, No 1, pp. 75-88

2) Abstract

The candidate's PF1 application form, which was approved by the Research Degrees Sub-Committee of Sheffield Hallam University in February 2004, included the following statement in the Abstract:

The United Kingdom's population censuses have been a vital source of information for the formulation of policy. This application of census data has been for spatial classifications from national down to sub-local authority levels. Traditionally the information from the censuses has been primarily in the form of fixed, pre-determined tabular output. This has restricted the usefulness of the censuses for policy-making processes.

The 1991 and 2001 censuses have provided data in a new format (microdata) which potentially overcomes many of the constraints inherent in earlier censuses. The sets of microdata are large samples of anonymised records (SARs) which contain over one million records of individual persons.

The research undertaken by the candidate examined whether, and how, these sets of microdata could be successfully utilised by urban and regional organisations to improve the quality of information available for the formulation of their policies. In

principle the sample sizes and ranges of variables available in the SARs, used in conjunction with appropriate analytical techniques, could provide a powerful mechanism to improve policy formulation. In practice a set of developments would need to be undertaken to maximise the effective utilisation of microdata for policy applications. The candidate has published a series of refereed research papers which have explored and developed some of the main aspects of this field of study. The research work required the identification, development and application of techniques which would be both valid with the data available and which would be simultaneously operationally useful to the policy-making agencies. In this context it is argued that the research has made original contributions to knowledge.

3) Critical Appraisal

Introduction

This application for a doctorate on the basis of published work rests upon the set of papers listed in section one of this submission. These have been published over the time period from 1996 to 2002.

This critical appraisal follows the procedures and guide notes issued by Sheffield Hallam University. The first section explains the overall aims of the research programme. This includes a necessary discussion of the context in which the research was initiated. The research programme is then described and this is followed by an analysis of the component parts. A detailed synthesis of the research as a coherent study is then presented. The final section provides an assessment of the extent to which the works have made a significant and original independent contribution to knowledge in the field. In addition to these sections, as defined by Sheffield Hallam University's procedures, a further section is then provided which identifies some areas with further research potential. The candidate is already developing a

Aims and Contextual Background

To understand how the aims of the research were developed into a coherent framework and provided a structured research programme it is necessary to summarise the candidate's experience of using data from earlier censuses, and to state some conclusions drawn from that work.

From 1969 to 1974 the candidate worked in the Operational Research Unit of Teesside County Borough Council and then in the Research and Intelligence Unit for Cleveland County. Operational Research at Teesside was primarily charged with developing computerised systems for improving decision-making within the policy processes of local government. Whilst employed at Teesside the candidate was involved with a joint research project between Teesside and IBM's national research centre at Peterlee. The project resulted in the research and development of a prototype computerised long range financial planning model for strategic and structure planning (Gardiner and Ward, 1974; 1980). This used census data and it rapidly became apparent that one analytical limitation was the inflexibility of the census data set as the output was in the form of pre-determined tables. The

requirement for a more adaptive output response had therefore been identified. Similar problems were encountered by Gardiner and Henneberry (1988; 1991) when attempting to undertake background research for studies of regional property markets.

Following the release of the 1981 census data a user's handbook was published in 1983 (Rhind, ed.). This included a chapter which summarised the position at that time concerning the release of microdata from the British census, and reviewed the arguments for and against such an action. This raised issues similar to some of those which had emerged in the Teesside research. The handbook noted that the Office of Population, Censuses and Surveys (OPCS) and the central government were not implacably opposed to the release of microdata and had asked potential users for submissions about the form which such data might take.

In 1990 the UK Census Offices agreed in principle to release microdata samples from the 1991 census. This was after a working party from the Economic and Social Research Council (ESRC) had developed the case for 'Samples of Anonymised Records' (SARs). This was based on research which had determined that there was an acceptably low risk of identification of individual persons or

12

households from the census data files. Consequently the ESRC agreed funding for the release of sets of microdata from the UK 1991 census for academic research purposes. Two sets of data were eventually released - the 2% sample of individuals (2% SAR) and the 1% sample of households (1% SAR). These were (in current 2004 ESRC terminology) large data sets - the 2% SAR, for example, consisted of approximately 1,100,000 records of individuals, and eventually provided information on over 60 different characteristics (including derived variables) of those individuals.

As the SARS are at the individual microdata level it is possible to transform data and undertake analyses which are either not possible or are not valid with the published standard census tabulations. These issues and related implications were discussed in detail in The 1991 Census User's Guide (Dale and Marsh, 1993). It was possible, for example, to extend cross tabulations to include several control variables of interest to policy-makers, subject only to limitations on cell sizes and statistical validity. Also it became possible to develop multivariate analyses which potentially could overcome these problems and simultaneously

lead to more sophisticated models of social, economic and demographic structures of central importance to policy-making.

One specific technical question of potential importance was linked to the ecological fallacy problem, which is concerned with drawing inferences regarding individuals from aggregated data, as originally propounded by Robinson (1950). This states that it is fallacious that a correlation at one level of aggregation will apply at another level and, in particular in the present context, that a correlation at any level of aggregation will not necessarily apply at the level of the individual. In applied policy work using census data the problem occurs when two or more separately aggregated sets of data (from standard census tables) are used to undertake correlation, etc., as part of, for example, investigations into the establishment of area-based policies, which become established on the basis of the apparent existence of spatially concentrated population characteristics (such as multiple deprivation). The availability of the 2% SAR clearly allows this ambiguity to be overcome since, by definition, this set of microdata is at the individual level and is not aggregated to area level(s). The ecological fallacy problem is therefore eliminated at the individual level and the 2% SAR provides policy-makers with a much clearer

statement of the actual level of link between characteristics of the target population within defined spatial areas. With this ambiguity removed it becomes more meaningful to use census microdata to provide a range of information to the policy-making process, such as benchmarking targets linked to key indicators, more precise targeting of specific sub-groups of the population within spatial areas identified for resource allocation or policy action, comparative analyses across spatial areas, and to facilitate in-depth analyses, surveys and related research pointing the way to further development of information sources and additional indicators at relevant spatial levels reflecting specific socio-economic and demographic features of policy interest to the organisations concerned.

The candidate recognised that the ESRC initiative could provide a powerful mechanism for handling such problems, identified originally at Teesside, and more generally throughout the field of central and local government. Consequently the candidate began to experiment with these data sets, once they became available to the academic community in 1993/1994. Results from initial investigations were very encouraging. The candidate therefore

concluded that a research programme using the SARs would be viable.

The philosophical basis of the research programme related to operational research and management science approaches to decision-making and to econometric modelling and statistical techniques which were used by Teesside and Cleveland County local authorities. In principle operational research techniques are based on the application of the scientific method to the solution of organisational problems. Essentially, therefore, this places the candidate's research programme within a rational positivist framework. However the approach was also heavily influenced by the candidate's working experience at Teesside and Cleveland County, which had demonstrated that straight forward application of econometric, statistical and operational research methods would often fail to be successfully implemented unless effective communication (of the concepts and results to senior managers and policy-makers) had been undertaken. The candidate noted that ineffective implementation could often be traced to problems which were not of a technical or analytical nature. These barriers often appeared to have political (organisational and/or party)

origins, or were linked to other traditional policy-making practices and procedures.

On the basis of these points it is argued that an alternative, and contemporary classification, of the candidate's research programme would place it within the evidence-based category of governments' approaches to policy-making. This classification has been discussed by Davies, building on his experience working within the Prime Minister's Strategy Unit (Davies, 2004). A commitment to evidence-based policy has been part of the modernisation and reform of the machinery of government by the Blair administrations. Evidence-based policy within the UK political system has been defined by Davies as an approach which puts the best available evidence from research at the heart of policy development and implementation in order that people could make well informed decisions about policies, programmes and projects. The central theme of this doctoral submission, which is concerned with how census microdata can be used to improve the quality of information available for formulation of urban and regional policies, therefore relates very closely to the key tenet of evidence-based policy. Furthermore Davies also identifies and describes different categories of research evidence used by policy-makers. These

categories include economic/econometric evidence, statistical modelling and descriptive analytical evidence, which are the main quantitative techniques used in the candidate's research programme.

However, evidence-based policy also recognises the importance of other types of research evidence in policy-making processes. These include, for example, impact and implementation evidence, and also evidence from attitudinal studies and consideration of ethical issues. The candidate acknowledges the potential importance of these alternative sources of research evidence, and considers that there are opportunities here for the development of further programmes which research ways of combining census microdata with other forms of evidence. These would further extend the current interest in quantitative data fusion techniques to an examination of methods of integrating census microdata with qualitative evidence. Effective implementation of policy often relies on qualitative data from in-depth interviews, focus groups and other methods of consultation. Research could be undertaken, for example, into how microdata based empirical evidence could feed into these procedures, how the output from these methods could in turn be used to refine further microdata based analyses, and

hence provide an assessment of the potential for overall improvement in the implementation of evidence-based policy-making. This approach would represent the development of proposals already contained in some of the PF1 list of papers to use microdata results, for example, to refine socio-economic surveys which would be commissioned by policy-making organisations. However, such an approach would represent a major extension and change of focus compared to the work described in this doctoral submission, which concentrated on the quantitative analysis of census microdata. Such an extension would provide the basis for a further research programme, separate from, but building on, this submission.

Finally, in this context, it is also appropriate to note that a range of factors influence policy-making in government. These include the key factor of evidence, but also such factors as experience, expertise, judgement, values and traditions, etc., which all play important roles in policy processes. The candidate would concur with this view of policy-making within organisations, and would argue that this view is consistent with earlier comments about the political barriers impeding the implementation of projects in Teesside and Cleveland. It is accepted that a research programme

based on census microdata needs to acknowledge that these factors may generate operational constraints in some policy-making situations, and thereby impact on the effectiveness of implementation of census based policy-making. This is, in part, the reason for the emphasis in several of the PF1 list of papers on the need for effective communication of census results and information to policy-makers. However, it is again argued that the central aims of the candidate's research programme were best achieved by concentrating on the technical and analytical aspects of using census microdata in public policy-making. Developing other effective ways to encourage policy-makers to use appropriate evidence involves research on how to change human behaviour within organisational structures, and is a major challenge requiring further research programmes based in social and behavioural science.

The aims of the candidate's research programme were therefore based on these conceptual and methodological issues and on the contextual background already described. Consequently they formed the basis of the statement in the Abstract and provided the hierarchy of aims outlined below.

The main aim was to determine whether the sets of microdata could be successfully utilised to improve the quality of information available for the formulation of policy by organisations dealing with urban and regional affairs. This linked to a second inter-related aim which was to determine in specific detail how this use of the microdata sets could be accomplished.

These central aims in turn relied on answering a set of questions linked to a series of second level aims. These were primarily related to technical and modelling problems including:

i)   identification of appropriate areas of policy and related indicators;

ii)   identification and development of appropriate analytical methods;

iii) operationalisation and assessment of these methods, including the impact of technical issues such as the 'ecological fallacy' problem, multicollinearity, etc., and the more generalised problem of developing methods which provide effective interpretation and communication of results to policy-makers.

The result of the subsequent work undertaken forms the central component of the published research items for this doctoral submission. However, before moving onto a detailed description of the research programme some of the issues raised in this section require further discussion.

First of all it should be noted that the research was restricted to the use of the 2% individual SAR. This was primarily to explore one of the main relative strengths of the 2% SAR, which is that this set of data has a local authority district level geography available for policy based analyses. For much of the research the 2% SAR provided a geographical level of analysis appropriate to much of local authority policy formulation (at overall local authority level). However, clearly a threshold of 120,000 population reduced the validity of the results for applied policy-making for some of the resultant combined smaller districts. The 1% household SAR, which has regional level geography built into the dataset, would have been potentially useful for only a limited subset of the analyses in, for example, Item 7. This explored Regional Development Agencies' strategies and policies, but even in this paper the discussion also drilled down to sub regional and local

authority level analysis, for which the individual level SAR was required. This question about the usefulness of the SARs' geography in the context of policy-making is discussed in more detail in several of the research papers, and in effect provides the main policy orientated reason for the sub-local authority level microdata proposals examined in stage 5. These formed the basis of the ESRC request to National Statistics for the release of small area microdata (SAM) from the 2001 census.

At this point it is also important to acknowledge that further sets of microdata were available for policy-making, but these were not considered or used in detail within this research programme. In part this was simply because the release of the 2% SAR presented the opportunity to assess a new set of microdata which, by definition, represented a potentially worthwhile research programme in its own right. Furthermore, the other main sets of microdata available such as the Longitudinal Study, General Household Survey and Health and Lifestyle Survey provided either smaller sample sizes, or incomplete geographical coverage, or confidentiality rules precluded the simultaneous disclosure of detailed geographical socio-economic and demographic output. Clearly these datasets can provide appropriate empirical evidence

for certain categories of policy-making, and therefore provide the base for a range of research programmes.

The 2% SAR was seen as providing a different set of research opportunities based on the availability of sub-regional and larger local authority geography in conjunction with the use of individual records, complete national coverage and a relatively large sample. The decision was made, therefore, to undertake a research programme, using these attributes of the 2% SAR, to examine how microdata from the UK population censuses could be used to improve policy-making. The main approach was seen to revolve round the application of primarily quantitative statistical and modelling techniques to the census microdata. However, it was recognised that many of the factors identified by evidence-based policy-making might be operative when attempting to implement SAR based research proposals. These factors (described earlier in the discussion about evidence-based policy-making) include staff experience and expertise, judgement and working traditions. Some of these factors were considered to be outside the remit of a research programme based on census microdata. However, others were implicitly, or explicitly, recognised in the listed publications, depending on the specific objectives of these papers. So, for

example, from an early point in the programme it became clear that financial resources and capacity constraints (primarily lack of skilled staff as well as computing limitations) in public sector organisations were placing severe limitations on the ability of these organisations to respond quickly and positively to the research opportunities presented by the sets of census microdata. This was seen as a spur to further academic research efforts to address such issues, and can be seen in a shift in the research emphasis from Item 3 onwards.

Finally it is important to note that for this doctoral submission the ecological fallacy problem became less important after 1996 because of the publication in that year of the research work by Fieldhouse and Tye (1996). The authors used individual level data from the 2% SAR in a study of deprivation to demonstrate that the use of area level data could be subject to the ecological fallacy. They showed that the strength of the relationships at the individual level were not as powerful as aggregate data suggested. They concluded that for policy-making at the local authority level it would be unwise to rely on a system of areal targeting or ring fencing, but that at "... a local level, however, such problems may not arise. The geographical areas used in this analysis are large local

authority areas (populations over 120,000) and were always very unlikely to identify homogeneous pockets of deprivation. The use of small-area statistics may allow the identification of deprived areas with much higher concentrations of deprived people than reported here (although an individual-level comparison is not possible at that scale). It is at the local authority level where areal targeting procedures based on aggregate level analyses of deprivation are most inappropriate." (Fieldhouse and Tye, 1996, page 257). The research conclusions of Fieldhouse and Tye meant that the candidate did not devote further research effort into the ecological fallacy problem as such. However, the resource allocation and targeting implications of these conclusions were very important, and reference was made to this problem in subsequent publications by the candidate where this was particularly appropriate in the context of policy-making processes.

Description of the research programme

The research programme has gradually evolved over the past decade. This was in part as a result of the experience gained from the initial research efforts and partly because of a series of developments pertaining to the microdata sets themselves. The programme eventually comprised four main stages. However there is no clear cut-off point for research in such an area, and it should be noted that a further series of developments resulted in an extension to the work, which in effect was a fifth stage. Further research stages into the future can also be envisaged. These are outlined later in this submission.

Table 1 below summarises the progression of the candidate's

research through the four main stages, plus the additional fifth

developmental stage.

Table 1: Summary of Key Stages in Research Programme:

Stage	Items relevant to Stage	Key Aspects or Tasks
1	None of the PFI working papers	Familiarisation with datasets and computer systems
2	1, 2 and 3	Used original structure of 2% SAR microdata; Limited use of spatial variables
3	4, 5 and 6	Development and use of explicit spatial analysis; Generation and use of specific policy defined derived variables; Links to other aggregated datasets and comparative evaluation of performance of 2% SAR with reference to these other sets of data
4	6, 7 and 8	Development of provisional proposals for the use of 2001 census microdata sets
5	9	Examination of the feasibility and case for a sub-local authority spatial level microdata set from the 2001 census

Stage 1 was essentially devised as an exploratory phase. Initially

the work was directed at familiarisation with the data set itself – by

definition nobody in the UK academic community had had any

experience of using this large data set. Stage 2 was envisaged as

an initial attempt to link the SARs to policy-making. At this point the methodology employed accepted that the conceptual basis and structure of the SARs was fixed and given. Applications were considered on the basis of using the variables in their original format (except for changing the number of categories). Areas of policy interest were identified and analyses were then undertaken which were appropriate to that policy area, but which were limited by the constraint imposed by the nature and inherent characteristics of the data set.

Stage 3 in the research programme followed a process of reflection about the approach adopted in the first two stages. One of the major problems identified from stage 2 was that the capability to include explicit analysis of a spatial variable was essential for many areas of policy in urban and regional policy studies. Secondly, an analysis could only be relevant if there was a near perfect match between the microdata variable(s) and the policy issue. Essentially what was required was an exploration of how to create, or derive, variables which were directly relevant to a specific policy under consideration.

This recognition of the limitations of the data set led to a realisation that further progress in this research area required identification and development of more specific analytical methods by which the SARs could be effectively utilised in the formulation of policy. The key mechanisms were seen to be:

a) to develop further the use of the SARs data at the individual level, but using derived variables (usually policy defined indicators);

b) to aggregate the SARs data to appropriate spatial levels, including derived variables, and then use in a second stage with other conventional aggregated urban and regional data sets;

c) using a) or b) as a first stage before analysis with the complete 100% census data on-line to the Office for National Statistics (ONS).

At this stage in the research several of the main methods by which the SARs could be used to improve policy-making had been

explored and discussed in the published papers. By this time the 1991 census data had already been released for several years and the detailed planning of the 2001 census was beginning to gather momentum. This shift in emphasis to the 2001 census effectively formed the underpinning and focus for stage 4 of the research.

Stage 4 of the programme examined how proposed forms of microdata and other customised output produced by ONS from the UK's 2001 census could be used to improve the quality of information available to policy-makers in urban and regional organisations.

Stage 5 represents research based on a proposal for a further spatial level at which 2001 census microdata would be made available. During the time period covered by stage 4 a research project was proposed by the Cathie Marsh Centre for Censuses and Survey Research at the University of Manchester. This project, funded by the ESRC, examined various aspects of proposals for the release of sub-local authority spatial level microdata from the UK 2001 census. As a result of the research work already undertaken by the candidate an invitation to join the ESRC team was extended to the candidate, initially as an advisor

on policy aspects. In essence this project formed the focus for

further development of the research.

Analysis of Component Parts

This part of the critical appraisal provides an analysis of the component parts of the programme. The discussion begins by identifying the type of publication and authorship represented by each item. A distinction is made between publications which appear in the PF1 list and a secondary list of works. These secondary items are identified because they represent interim reports, or they provided parallel discussions, which help to understand the reasoning underlying the PF1 set of publications. A third category (Further Works) is also presented. This category includes work which provides evidence demonstrating how the main body of research in this submission has been further developed.

A summary of the type of publication is provided in table 2 below:

Table 2: Type of Publication

Type of Publication	Frequency of Items: PF1 List	Frequency of Items: Secondary	Frequency of Items: Further Works
Refereed Papers	6	0	1
Editorially Refereed Papers	1	0	0
Working Papers	1	0	0
Conference and Occasional Papers	0	2	2
Workshop Papers	0	0	3
ESRC Research	0	0	1

The refereed papers have all been published in journals of high standing and reputation. Four of the publications listed on the PF1 submission appeared in journals which are considered to be of high international standing (*Urban Studies* and *Environment and Planning C*). A further two papers were published in *Local Government Studies* and *Planning Practice and Research,* both of which are regarded as of high national research reputation with importance attached to the applied practice element. One further paper appeared in a publication which uses editorial editing as a standard procedure, and is directed at the research and information systems group of practitioners in central and local

government. The ESRC funded research was a project co-ordinated and led by the University of Manchester.

The published work can also be analysed in terms of the degree to which the list of publications represents sole authorship or co-authorship. Six of the papers in the PF1 list were submitted under the sole authorship definition. Co-authorship in the PF1 list is defined to mean that the candidate and one other person were equally responsible for the publication. Typically in these cases the other author was primarily responsible for researching the theoretical and policy context and the candidate developed the empirical analysis. Both authors were then jointly responsible for writing up the research and providing an evaluative summary of the work. This category is represented by two papers. The "Further Works" heading includes a refereed paper (listed as Item 9 after the PF1 set of items), which is also co-authored.

It is also possible to categorise a list of publications by intention or purpose. For this research programme all of the publications concentrated on the use of census microdata to investigate areas of policy and improvements in the quality of information for policy-making. This was the main focus of the research. The fact that all

of the items link back to this common purpose emphasises the underlying coherence of the programme and that an attempt was made to maintain a consistent theme running through the research.

A further method of analysing a research programme would be to look at how the relative emphasis placed on aspects of the research approach varied from one publication to another. An examination of the publications in this research programme will confirm that the main emphasis in all these publications was a concentration on methodological issues and operational methods, using empirical results from selected case studies to identify and develop issues raised. The main differences, therefore, in the publications were in the degree of sophistication of spatial analysis undertaken (which is of crucial importance in the context of urban and regional policy-making), the size of the effective data set available for a specific application, and the inter-related level of complexity of analytical methods employed, together with the degree of precision in matching the techniques developed to policy processes and policies. In general, the relative importance of these changed and developed over the lifetime of the research

programme. These points are developed in greater detail in the

next stage of the critical appraisal.

Synthesis of the works as a coherent study

It will be argued that there is a clear coherent theme running through the research programme. By definition this coherence underpins the structure of the programme and the analysis of the component parts of the programme. The following discussion builds on those sections and, therefore, links back to some of the earlier discussion. However the emphasis will be placed on a description of the underlying elements which provided consistency and coherence to the research programme. In so doing, a degree of evaluation of the individual components will be undertaken and integrated into a reflective analysis of the overall programme. In turn, since these elements were inter-connected, by developing the discussion about them an effective synthesis of the research will be provided. An overview of the main ways by which the components integrated into a coherent programme is provided in table 3.

Table 3: Summary of key elements providing coherence

Item	Policy Context	Spatial Scale	Policy Indicators Redefined	Use of Indicators
1	Transportation Studies; Links to groups at risk and elderly	Local Authority; Single SAR area	None redefined; Except for re-coding to reduce categories and creation of binary dependent variable	Cross tabulations and Significance tests; logistic regression for explanatory analysis
2	Transportation Studies; Cycling policies	Local Authority; 11 SAR areas	None redefined; Except for re-coding to reduce categories	Cross tabulations and identification of characteristics of population cycling; use in re-targeting of policy formulation
3	Standard Morbidity Ratios and local government finance	Local Authority; Single SAR area	None redefined; Except for re-coding to reduce categories and creation of binary dependent variable	Logistic regression; Model output used in comparison with Sheffield Central Policy Unit's results derived from conventional census output tables; Demonstration of further improvement

Table 3 (continued):

Item	Policy Context	Spatial Scale	Policy Indicators Redefined	Use of Indicators
4	Regeneration Studies	Local Authority; Single SAR area	Derivation of key policy defined variable	Cross tabulations with policy defined variable and preliminary comparison of output with other research using conventional census output tables
5	Regeneration Studies	Local Authority; 237 SAR areas	Derivation of key policy defined variable	Cross tabulations with policy defined variable and comparison of output with research using conventional census output tables; Logistic regression with explicit spatial variable (for the 237 SAR areas)
6	Standard Spending Assessment (SSA)	Local Education Authority (LEA); 237 SAR areas combined to LEA level	Derivation of sets of explanatory variables at individual level to link to SSA indicator	Aggregation of explanatory variables to LEA level; Statistical comparison of performance with official SSA formulae
7	Strategies and policies of Regional Development Agencies (RDAs);	9 English Regions; Local Authorities within the Yorkshire and the Humber Region	Derivation of RDA defined core indicators	Implicit cross tabulations with policy defined variable; Logistic regression with explicit regional spatial variable and interaction effects

Table 3 (continued)

Item	Policy Context	Spatial Scale	Policy Indicators Redefined	Use of Indicators
8	Housing Needs and Local Government Indicators; Assessment of potential value of 2001 census microdata	237 SAR areas; Local Authorities within the Yorkshire and Humber Region	Derivation of Housing Needs Indicators based on measures of overcrowding	Cross tabulations used in conjunction with classification techniques (GB Profiles) at enumeration district level; Assessment of potential usefulness of sub-local authority level microdata using statistically defined area classification
9	Potential use of Small Area Microdata (SAM) from 2001 census; Limiting long term illness case study	SAM data for 7 local authority areas provided as a special file by the Office of National Statistics. Local authorities contained between 1 and 15 SAM areas	None	Cross tabulations (actual tables were not allowed to be published because of disclosure risks); Logistic regression models with a sub-local authority level spatial variable included

Within the research programme outlined in table 3 the items represented, in general, increasingly sophisticated levels of modelling, including spatial analysis, and/or closer matching of

indicators to specific policy requirements. The key requirement, and therefore coherence of the programme, remained the need to determine how census microdata could be utilised by urban and regional organisations to improve the quality of information available for the formulation of their policies.

Items 1 and 2 used policy and explanatory indicators in the original non-modified 2% SAR based format. The spatial level increased from a single SAR area in Item 1 to eleven areas for Item 2. However, the level of statistical analysis undertaken was constrained as a result. The main emphasis in these two papers was concerned with estimation and significance of explanatory socio-economic and demographic models. Item 3 began the process of more specific focussing on policy requirements in local authorities and similar organisations. However, the main policy defined indicators remained in the original SAR format. Items 4 and 5 were linked, and represent a move towards the inclusion of an explicit spatial variable at local authority level. In addition a second major change in the methodology was the creation of SAR based derived policy variables. The research then examined, in Item 6, how these derived policy indicators and spatial variables could be aggregated and then used in combination with other

urban and regional sets of data. At this point in time detailed debate about the nature and content of the 2001 census output was underway. It was appropriate, therefore, to build on the existing platform of research expertise to assess the opportunities which the release of the 2001 data would offer. This work was contained in Items 6 (to some extent) and Items 7 and 8. The latter paper also began an examination of sub-local authority level spatial variation, and the implication for targeting of resources on sub groups of the population within specified spatial areas. However, this analysis was constrained to the use of statistical based classification of small areas. The final item, which could not be formally included in the PF1 list, discussed some preliminary results from research using actual small area microdata. A detailed discussion of these points, cross-linked to the main stages of the research programme, is presented next.

As indicated in the earlier outline of the research programme, the first stage of the research involved an exploratory phase which required the candidate gaining experience in the use of the SAR data sets. Although the computer and information systems employed do not constitute a central aspect of the research programme as such, nevertheless some comment is appropriate.

This is because the technology available at a given time period constrains the levels of analyses that can be undertaken, and determines, in part, the quality of research which is possible. In the case of the census microdata this was a particularly important consideration because of the large size of the data set and the storage location and arrangements for access to the data.

At the start of the research programme the data was held on a Cray supercomputer running under the Unix system at the University of Manchester. Transfer of data and output to mini-computer systems at Sheffield Hallam University was sometimes disrupted by the poor reliability of the (relatively) rudimentary communications technology. Therefore, in the initial stages of the research programme the solution adopted was to transfer a component of the full microdata set (for, say, one local authority district or, alternatively, a limited number of variables for a large set of districts) to Sheffield for further detailed analyses to be undertaken. However, after these initial teething problems had been overcome, and with rapid improvements in communication and transmission of data and results, it became possible to use the full computational power of the Cray and develop further the research at hand.

The second stage of the research programme is covered by three items. The first two were co-authored published papers with Hill. The two items are analysed separately, but there are common links between these papers. The third item was a sole authored published paper by the candidate.

Item 1 is a paper which was one of the first to be published using the 1991 UK census of population SARs and was an early demonstration that it was possible to link effectively the analysis of microdata to an aspect of urban and regional policy. It is worth recording parts of the conclusion to this paper, because it encapsulates some of the central core discussions underlying the candidate's research programme. The conclusion stated that:

> "The data set has been shown to be appropriate for the exploration of social issues in transport and for comparative study in differing UK policy contexts and environments, from the regional down to the scale of most local planning and transport authorities …….. The empirical results have established the strength of the relationship between a number of demographic and socio-economic variables and access to cars among the elderly" (pp 278-279).

The paper also concluded that quantification of differentials in access, because of demographic and socio-economic differences, would enable their impact to be addressed by the precise targeting

of transport policies, provisions and concessions. It is interesting to note that, although these statements were made before the election of the present Government, they are relevant in the policy context of placing emphasis on targeting specific sub-groups of the population in specific spatial areas, which has been a strategy promoted by this Government.

The work also provided an exploratory insight into the detailed technical and statistical advantages and disadvantages of samples of microdata. These are noted in this first paper, and they form the basis of some of the areas for development in successive papers. The problem of sample size and small cell frequencies (in cross tabulation and significance tests) is raised on page 274, followed on page 275 by the need to cascade categories. Related problems are discussed in the context of the results for the logistic regression models, which exhibit symptoms of multicollinearity. This problem was identified as being due in part to the reliance on dichotomous variables necessitated by the relatively limited data set for the Sheffield case study[i]. The paper anticipated that the problem might be less severe with a larger sample of data which would not require as many restrictions on the number of categories.

To utilise a larger sample of data would require explicit recognition and use of the spatial variables in the data set. These variables, which provided identification of regions and local authorities, were needed for modelling the important spatial aspects of policy formulation (as well as providing potential benefits from increased sample size). The first paper had had an implicit acknowledgement of the importance of spatial considerations when it discussed (for example) comparative studies from regional down to local authority levels. Likewise the use of Sheffield for the case study in the first item, by definition, implicitly represented a choice of one spatial area, rather than some other area(s).

Item 2 began this examination of how spatial attributes of the data could be utilised. It undertook a set of analyses for a group of local authority areas, rather than concentrating on one (local authority) area - whilst simultaneously continuing to provide an analysis of the socio-economic variation in the sample of individuals. At the time of this research work the problems with the communication and information technology still represented a severe constraint on the number of spatial areas and size of the data set which could be handled simultaneously. For this study the solution was to use a

non-random selection of eleven SAR areas which were chosen on the basis of maintaining spatial compatibility with the areas used by other researchers in earlier work. The spatial selection process generated a sample of approximately 26,000 individuals for the travel to work variable, and a sub sample of over 1,300 individuals who cycled to work. The analyses reported in this publication were primarily based on cross tabulations. Once again the problems of sample size arose in some of the analyses, which again required the combination of some categories for some of the variables. Nevertheless, this selection of SAR areas (and therefore by definition of sample size) was deemed to be an acceptable solution for this specific case study (within the perspective of the wider range of research already undertaken in this field). However, it was becoming clear that this problem might be a significant constraint in other areas of policy.

With the benefit of hindsight it can also be seen that the emphasis in each of these papers was directed towards the estimation process and determination of the significance of explanatory socio-economic and demographic variables linked to a dependent variable – even though the dependent variable in each paper was, indeed, a policy based indicator. Although the second paper does

specifically identify the concept of using microdata to generate baseline data for policy, nevertheless it could be argued in retrospect that the emphasis on the policy implications needed ideally to be made more precise and explicit.

Item 3 is a paper which developed from discussions with members of the Central Policy Unit of Sheffield City Council[ii]. The aim of the research reported in Item 3 follows on from the points made above. It represents an effort to focus the work more closely on the central research questions of the programme. An attempt was made, therefore, to demonstrate that SAR based analysis could:

a) effectively replicate analysis required for a specific policy or policies and

b) effectively improve the quality of the analysis and so improve the quality of the information available to policy-makers.

The article, which appeared in the British Urban and Regional Information Systems Association (BURISA) newsletter, was written with the intention of reaching an audience of practitioners in local authorities and similar organisations. It was based on developing a methodology derived from an earlier article by Smith and West

(1993) from Sheffield City Council's Central Policy Unit which used

conventional 1991 census data for Sheffield to advance arguments

for changes in the central government's Standard Spending

Assessment (SSA) formulae. Their work revolved round the

establishment of a case for introducing a Standard Morbidity Rate

from the 1991 census based on the limiting long term illness (LLTI)

indicator available (for the first time) from the census. Using the

SAR data for Sheffield it was possible to undertake equivalent

analyses and compare the SAR based results with those of Smith

and West. However, the article also demonstrated how it was

possible to extend their work by using logistic regression

techniques which allowed for the simultaneous inclusion of all the

identified explanatory variables, and which also avoided any

possibility of the ecological fallacy problem. It is worth noting that

publication of this paper coincided with that of the paper by

Fieldhouse and Tye (1996), discussed earlier in this submission,

which investigated the actual importance of the ecological fallacy

problem using empirical evidence from the SAR dataset.

At the time of publication the case for the use of the SARs

appeared to be clear-cut in the case of this specific application.

This is still essentially the case. However, as a result of later work

undertaken by the candidate one caveat should now be noted. The work by Smith and West used variability in socio-economic and demographic data at ward and enumeration district levels to investigate corresponding spatial variations in morbidity. The SAR data used individual level differences in personal attributes to model the same incidence of LLTI. Identification of spatial variability was not an issue in its own right in this example. However, for some areas of academic and applied work at sub-local authority spatial levels, variability is the central focus of the analysis – this issue is addressed later.

It is also important to raise one particular point for discussion at this stage. This is the issue of effective communication of information to the user community. It rapidly became clear that, perhaps not surprisingly, the forms of SAR based analysis which potentially provided the greatest insight and understanding were those structured around the most sophisticated techniques. How these results could be set in a format which was clearly and easily understood is also a point raised later.

The third stage of the research followed and developed from the experience gained from the second stage of the programme. This

third stage consisted of two published papers (both sole authorship), which will be discussed in detail, although some preliminary work was also undertaken and this is summarised first.

One of the key developments was to generate, using the SARs data at the individual level, derived variables specifically linked to policy defined indicators. As a first attempt in this direction a working paper was written which described this process for one spatial area (Sheffield). This working paper is included as Item 4. A second crucial development involved extending the analytical models to include a spatial variable explicitly, which is a central concern to spatially based urban and regional organisations. This development was driven primarily by these perceived requirements of policy – for example the need to demonstrate that the SARs could be used for various forms of comparative analysis, for performance indicators across regions or local authorities, etc. Fortunately this development also corresponded with significant improvements in the computing systems and communication links to Manchester, which meant that more complicated and computationally demanding analyses could be successfully undertaken. As a first step in this direction a conference paper was presented to the Academic Planners Conference at the University

of Nottingham in 1996 (unpublished). This paper extended the SAR based models from one spatial area (Sheffield) to two areas (Sheffield and Nottingham).

The experience gained from these papers fed directly into Item 5. This item, therefore, represented research which combined elements from both of these developments. Firstly, the paper represented an important extension in the spatial level of analysis. Much of the analysis and corresponding discussion was based on tables and models which included data from all 237 English 2% SAR areas. In this paper spatial variation was explicitly modelled by the inclusion of a variable which represented each SAR area as a category within the area variable, and therefore provided empirical estimates of the different effects of each area.

Secondly, for this paper a case study from the area of regeneration was used which identified a specific example of a policy indicator which was not directly available from the standard census cross tabulations. The paper demonstrated that it was feasible to identify and define an equivalent SAR based indicator. It proceeded to derive this SAR based policy indicator and then described how refinements could be made to this indicator, which would offer

further potential improvements in the targeting of policy initiatives towards specific sub-groups within the population.

Item 6 represents an area of research which concentrated on a further development of the analytical methods employed. In conventional analysis for policy it is often the case that two or more variables are used to model some aspect of policy. One important example of this has been the central government's Standard Spending Assessment (SSA) exercise for the distribution of central government funds to local authorities. This was based on a series of indexes, which in turn were based on formulae derived using ordinary least squares (OLS) multiple regression.

For this research one specific area within the SSA was used as a case study. The area chosen concerned the Additional Education Needs (AEN) index. The Department of the Environment's (DoE's) AEN index was based on an OLS model using variables already aggregated at local education authority level. These variables related to levels of lone parents, income support and ethnic minorities.

In this paper the SARs were used to select individual records of people with *all* of the characteristics which placed them in a 'vulnerable' group, as defined by the DoE's Index of Local Conditions (Deprivation). The criteria for selection of individuals were ethnicity, lone parent households, local authority tenure and being of school age. In effect this generated a new (or derived) explanatory variable which, by definition, represented that sub-group of individuals and could be used very effectively for targeting and resource allocation. This variable could then be aggregated to the appropriate spatial level (local education authorities) and then linked to a specific policy indicator, or general area of policy (the AEN formula in this case study).

The results of this innovation in methodology were reported in the paper. It is important to note that in this specific study it was possible to compare the quality of the SAR based procedures with a clearly defined, existing DoE methodology. In summary it was demonstrated that use of the SARs could lead to improved statistical results, as recorded by the coefficient of determination, which was one of the measures used by the DoE in the official SSA handbook.

In Item 6 the approach of the 2001 census was given detailed consideration and emphasis for the first time. The section in this item on 'Future developments and the implications for the 2001 SARs' (p 555) can be seen as signalling a significant shift in the main focus for research.

This specific shift in focus in Item 6 was at a time period when the 1991 census data had already been extensively researched and detailed planning of the 2001 census output was under consideration. As a consequence of the candidate's published work at that stage he was invited to join the Office for National Statistics (ONS) subgroup of the 2001 Census Outputs Working Party. This had a remit to investigate user requirements for SARs from the 2001 census. It provided an opportunity to contribute to the debate about the structure of the 2001 microdata, and an opportunity to talk to other academics and to ONS staff about their perceptions of the likely advantages and disadvantages of the 2001 microdata. In parallel to this process the candidate began to draw together the main strands from earlier papers and reflect upon the knowledge gained and conclusions to be drawn from investigating the 1991 microdata.

A further consequence of the candidate's publication record was that the candidate was invited to present a paper at the "How can census microdata help local government policy-making?" conference run by the Cathie Marsh Centre for Census and Survey Research, University of Manchester (1997 unpublished). This again provided useful feedback from practitioners in local government. Much of this work and subsequent debate from these meetings and conferences forms the basis and underpinning for stage 4 of the research. This focused on the question of how information for policy-making could be improved using microdata from the 2001 UK census of population. Discussion around this point is represented by two further papers described next.

As noted above Item 6 represented a major shift in the focus of the research programme, beginning an examination of how specifically the 2001 microdata could be utilised to achieve the objective of improved information. This research theme continued for Item 7. Once again the approach was to base the discussion around a specific case study which in this paper was linked to an examination of the strategies and policies of Regional Development Agencies (RDAs). Of necessity this examination was still based on 1991 data, but the analysis was used to identify how

proposals for the 2001 census could improve information in comparison with the situation existing with the 1991 data. This included discussion about linking samples of microdata and a second stage use of customised output, which would be provided by ONS.

The research presented in Item 8 followed on from that for Item 7. The article again used a policy study (in this case from the area of housing) to demonstrate how microdata from the 2001 census could be used (again if necessary in conjunction with customised output) for policy-making and resource allocation, primarily at local government level. However an important development and difference from Item 7 was the focus in the discussion on the level of the spatial dimension provided by the data set. The case study demonstrated the potential importance of data at sub-local authority spatial level. The research methods used in Item 8 employed a SAR variable which provided an enumeration district classification of residential area for each individual person. This classification was based on a statistical definition of spatial areas. Such an approach is appropriate for the type of policy described in Item 8. This is because identification of exact spatial location is not required - just the classification of each spatial area. In such a

policy-making environment the paper was able to demonstrate that microdata, perhaps in conjunction with customised output, could provide operationally useful information for policy. In so doing, the study also indicated that spatial location at sub-local authority district level might be potentially an important variable in explaining variation in the incidence of such policy indicators.

However, local government and similar organisations also require actual spatial location information for a wider and more extensive range of policies. The paper noted this, with the implication that the 1991 SARs could not assist in such areas of policy since actual detailed small area data could not be derived from the microdata from that census. The concluding section identified and discussed potential mechanisms using the 2001 census to overcome such problems. Some of the comments made in that conclusion were linked to research being undertaken by the candidate in conjunction with the Cathie Marsh Centre for Census and Survey Research at the University of Manchester. The background and outcome of this research is described next.

At this point several overlapping strands of potential areas of research development had been pulled together by a decision by

the Cathie Marsh Centre to apply for ESRC funding for a project under the Census Development Programme. This proposal was in part a result of the work which had been undertaken by a research team at Manchester (Dale and Elliot,1998) which had looked at disclosure risks from proposed 2001 SARs. Also relevant was a working paper (Brown and Dale,1998) which had provided an analysis of a survey of SAR users and their requirements for 2001 SARs. A series of consultation workshops entitled "SARS 2001" had also been held in 1999 to ascertain in more detail academics' and practitioners' views on the 2001 census microdata. The candidate gave presentations entitled "The case for a third SAR from the local authority perspective" (unpublished) at two of these meetings, one at the University of Manchester and the other at the London Research Centre. Further informal views were input to the process by some of the other researchers experienced in using the 1991 microdata.

The successful outcome of the ESRC application led to the formation of a project team for the period 1999/2000. The project's title was "SAR developments for 2001: Meeting user needs for small area microdata" (ESRC Award Number H507255161). The candidate was initially invited to join this research project as an

adviser on the policy implications of the proposals for the 2001 microdata.

This area of work, concerning the application of small area microdata (SAM) to the formulation of policy, constitutes the potential basis for a long term programme of further research leading up to the next census in 2011. The links to initiatives such as Neighbourhood Renewal and evidence based policy formulation are very clear. In the shorter term, within the time span of the ESRC project itself, several conference papers and provisional research conclusions were produced, and this work is summarised next.

In June 2000 a paper was presented at the "Census of Population: 2001 and Beyond" Conference, University of Manchester, by the ESRC project team. The paper was attributed to Brown M, Dale A, Elliot M, Fieldhouse, Pickles A, and Tranmer M, (from the University of Manchester), with acknowledgements to the consultants Martin D (University of Southampton), Gardiner C (Sheffield Hallam University), and Steel D (University of Wollongong). This paper was entitled "Microdata for Small Areas: the case for a third SAR" and was published on the web site of the

Cathie Marsh Centre for Censuses and Surveys. A further version of this research was also produced as an occasional paper. This was entitled "Microdata for Small Areas", and is also on the web site of the Cathie Marsh Centre as Occasional Paper Number 20. The candidate's contribution also included a paper for a "SAMs for 2001 Census" consultation workshop held at the Greater London Authority in 2000. The paper was entitled "Small area microdata from the 2001 Census: The local authority perspective" (unpublished).

In addition to these conference and workshop papers, two further refereed papers have been accepted for publication. One paper will be published in the Royal Statistical Society's Journal (A) (Tranmer, *et al.*, forthcoming). The second paper has been included as Item 9 after the PF1 list of published works submitted for this doctorate. This item discusses how the release of small area microdata from the 2001 UK population census could help to provide better information in the policy-making process at sub-local authority spatial levels. The empirical results provided estimates of the effects of area at ward level and separated these effects from those due to socio-economic and demographic characteristics of the population. The paper concluded that the small area microdata

results appeared to differ from results derived from tabular area output, which could have implications for the design and targeting of policies at sub local authority level.

Conclusion

The preceding discussion has provided a synthesis of the research programme and has established that this was based on a single coherent theme. Within this programme the stages represented different, and increasingly sophisticated, levels of analysis as the research developed over time, working towards the aim of determining whether, and how, sets of census microdata could be successfully utilised by urban and regional organisations to improve the quality of information available for the formulation of their policies. The main stages have been described in detail and consisted of a preliminary phase directed at essential familiarisation with the data set itself, followed by research studies into how simple SAR based analysis could be used for policy with non-modified variables. Working still towards the same set of aims the research developed and brought into consideration explicit utilisation of a spatial area variable in conjunction with a methodology based on the creation of derived variables. The research then continued using the newly derived policy indicator

and spatial variable with either the original microdata set, or after appropriate spatial aggregation of the data, in combination with other urban and regional sets of data. As already discussed, at this point in the research several of the main methods by which the SARs could be used to improve policy-making had been explored. Discussion about the output from the 2001 census was beginning to gather pace. Building on the experience gained from using the 1991 data the candidate began to direct the research effort towards assessment of the likely implications and opportunities which the release of the 2001 data would offer. This represented the next stage in the programme, which again was consistent with the underlying set of aims. This fourth stage represents the final stage possible to discuss fully in this application (under the existing University regulations). However some preliminary results for the SAM based research undertaken with colleagues at Manchester were presented in the final part of the discussion.

The significant and original independent contribution of the works to knowledge

Under the University's regulations the candidate is required to identify the extent to which the submitted works are the candidate's original independent contribution to knowledge. All of the works for this submission, with the exception of the first two papers, represent research undertaken by the candidate alone. Both publication one and two involved full collaboration and inter-linked work and research with the other co-author, R Hill. As stated in the PF1 document R Hill took primary responsibility for researching the contextual and conceptual background and information, and the candidate was primarily responsible for much of the analysis, interpretation and statistical evaluation of results. Joint responsibility was assumed for overall evaluation of the studies, and for writing and editing of the resultant papers.

A second requirement is to establish the significant contribution of the works to knowledge in the field. Some retrospective assessment is possible. The invitation to join the Manchester led ESRC project came about because of Manchester's recognition of

the quality and potential significance of the work already undertaken and published by the candidate. It should be noted and emphasised that initially the invitation to join the ESRC project was specifically to act as an adviser on the policy implications of proposals for the 2001 census microdata. Under the ESRC's official outcome evaluation grading system, this project was graded as "Outstanding: High quality research making an important contribution to the development of the subject. An Outstanding grade indicates that a project has fully met its objectives and has provided an exceptional research contribution well above average or very high in relation to the level of award", letter dated 29th August 2001.

Likewise the invitation to join the ONS working group resulted from recognition of the quality of the publications produced by the candidate. Similar comments also apply to the invitations from Manchester University to present papers at a series of conferences and workshops.

It is also worth emphasising that, as already noted, several of the papers in the PF1 list were published in journals with international research status (for example *Urban Studies* and *Environment and*

Planning C: Government and Policy). *Local Government Studies* and *Planning Practice and Research* would be regarded at least at national research status and are important for dialogue and dissemination of ideas to practitioners. By definition, publications at this level are only published after an extensive and exhaustive examination by referees, which is a measure in its own right of the significance and originality of the research embodied in these publications.

In the longer term it should also be possible for the second requirement to be judged by the actions of organisations within the urban and regional policy-making framework. However, by definition, this evaluation will to a large extent only become meaningful from approximately 2005 onwards when it will be possible to begin to examine the level and effectiveness of actual applications of the 2001 microdata and customised output in areas of policy-making.

Other evidence (but non-research based) is also available. As a result of the candidate's research and related publications record the candidate was invited to serve as a member of the steering

group for one of the clusters in the Distributed National Electronic Resource (DNER) programme run by the Joint Information Systems Committee (JISC) for all the UK Universities. This was funded by the Higher Education Funding Councils, which have given "JISC the remit to converge new learning environments with digital library developments".

The cluster (of which the candidate was a member) was concerned specifically with enhancing JISC data services for teaching and learning. The cluster steering group had responsibility for projects concerned with applications of the UK population census, business and economic analysis, etc. This included, for example, the census project "Developing the Collection of Historical and Contemporary Census Data and Related Materials (CHCC) into a Major Learning and Teaching Resource".

Finally it is important to reflect on the current situation with respect to the census microdata, and to look at some future possibilities for the potential usage of the 2001 SARs. It is relatively straightforward to identify some important areas for further work in this field, based on components of the research programme explained in this submission. One specific possibility relates back

to one of the aims of the research which concerns the need for effective communication of information to user communities. The empirical results provided by multivariate models are complex, and there are important issues to be addressed here about interpretation and communication of these results to mainstream policy-makers. One area for further research would be centred around the development of predictive models (linked to a specified area of policy). Such a possibility has been raised by, for example, Bramley and colleagues (1999) and comments by the candidate in Item 8 (p 84). The central analysis and output for policy would still be based on the same multivariate models - with all the technical advantages these provide. However, what is envisaged here would be a further step in the modelling analysis which would provide indicators showing relative levels (performance) by spatial area. The objective would be to generate a class of indicators which would be relatively easy to understand and interpret in the policy context (while retaining the analytical power of the multivariate models).

A related area will be the opportunity for comparative analysis using the 1991 and 2001 census data. Whilst this can be undertaken, of course, using standard output, once again the

advantages of the flexibility of microdata output would allow more specific comparisons to be drawn. One major methodological point to develop here would be that if multivariate analyses were used this would require comparison of (for example) logistic regression coefficients for specific spatial areas for 1991 and 2001. This requires refinement of standard econometric tests.

More generally, at the time of submission, ONS has agreed in principle to the release of a SAM set. However detailed specifications have still not been published. The spatial level proposed should be lower than the 2% 1991 SAR (currently the geography will allow for each local authority to be represented separately, with the exceptions of the Isles of Scilly and the Cities of London/Westminster), but will not go down to ward level data (as in the original ESRC request). Nevertheless the lower level spatial definition should facilitate a wide range of fascinating academic research and policy applications.

Furthermore, as from the autumn of 2004 National Statistics will provide a 'Controlled Access Microdata System (CAMS)' which allows bona fide researchers supervised access to a fuller microdata set, with more categories retained in key variables, and

at local authority spatial levels. This will be available for academic researchers, local and central government staff, and private sector users. Initially this will only be available in a secure environment in the National Statistics offices at Titchfield, but plans are in place to extend the facility to other regionally based offices of National Statistics. Further opportunities exist to enhance the quality of the information which can be extracted from the 2001 census. One possibility will be to take this approach one extra stage and use the empirical results obtained to form the basis of a request for further detailed commissioned output tables or multivariate analysis from National Statistics based on the full census database (this approach was suggested in several of the later papers listed in this submission) . Other opportunities also exist to enhance the quality of the information which can be gleaned from the census microdata for policy-making purposes. These include, for example, further development of the use of synthetic estimation techniques, and the combination of microdata with information from 100% tabular data. Further potential opportunities could evolve from, for example, the ESRC/JISC programme for the development of e-social science and in related applications of Grid technology, dependent upon future progress in those areas.

[i] However, it is important to put these comments within the context of current practice in, for example, local authorities – the analysis in this paper was based on a sample of approximately 3,500 individuals from the 2% individual SAR for Sheffield, which still represents a relatively large sample for much of local authority practice.

[ii] Although the publication date of item 3 is between that of item 1 and item 2, the work represented by item 3 was effectively undertaken after the completion of the first two items.

4) References

Bramley, G, Pawson, H, Satsangi, M, and Third, H, 1999, Local Housing Needs Assessment: A Review of Current Practice and the Need for Guidance, *Research Paper No.73, School of Planning and Housing, Edinburgh College of Art/Heriot-Watt University*

Brown, M and Dale, A, 1998, A survey of SAR users, their requirements for 2001 SARs and their views on dissemination and support", *Working Paper Number 6, The Cathie Marsh Centre for Censuses and Survey Research*
http://www.ccsr.ac.uk/publications/working/sarsuser.htm

Brown M, Dale, A, Elliot, M, Fieldhouse, E, Pickles, A, Tranmer, M, Martin, D, Gardiner, C, Steel, D, 2000, Microdata for Small Areas: the case for a third SAR, paper presented to the *"Census of Population: 2001 and Beyond" Conference, University of Manchester, http://www.ccsr.ac.uk/conference/papers.htm*

Brown, M, Dale, A, Elliot, M, Fieldhouse, E, Pickles, A, Tranmer, M, Martin, D, Gardiner, C, Steel, D, 2000, Microdata for Small

Areas, *Occasional Paper Number 20, The Cathie Marsh Centre for Censuses and Survey Research*

http://www.ccsr.ac.uk/sars/2001/Occ20.pdf

Dale, A and Marsh, C, 1993, The 1991 Census User's Guide, *HMSO*

Dale, A and Elliot, M ,1998, A report on the disclosure risk of proposals for SARs from the 2001 Census,

http://www.ccsr.ac.uk/publications/occasion/occ16.pdf

Davies, P, 2004, Is Evidence-based Government possible? , *Government Chief Social Researcher's Office, Prime Minister's Strategy Unit, Cabinet Office, London*

Fieldhouse, E, and Tye, R, 1996 Deprived people or deprived places? Exploring the ecological fallacy in studies of deprivation with the Samples of Anonymised Records, *Environment and Planning A, 28, pp 237-259*

Gardiner, C, and Henneberry, J, 1988 The Development of a

Simple Regional Office Rent Prediction Model, *Valuation*, 7, pp 36-

52 .

Gardiner, C , and Henneberry, J, 1991 Predicting Regional Office

Rents using Habit-Persistence Theories, *Property Valuation and*

Investment,9, pp 215-226.

Gardiner, C, and Ward, P W, 1974, A Long Range Financial

Planning Model for Teesside County Borough Council, *Operational*

Research Quarterly, 25, pp 55-64 .

Gardiner, C, and Ward, P W, 1980, A Long Range Financial

Planning Model for Teesside County Borough Council , in Collcutt ,

R H,(Ed.), *Successful Operational Research - A Selection for*

Managers, chapter 5 , pp 61-72 , Operational Research Society .

Rhind, D, (Ed.), 1983, *A Census User's Handbook*, Methuen

Robinson, W S, 1950, Ecological correlations and the behaviour of

individuals, *American Sociological Review,* 15, pp 351-357.

Smith, D and West, D, 1993, Developing the Concept of a Standard Morbidity Rate from the 1991 Census, *British Urban and Regional Information Systems Association, 111, pp. 4-8.*

Tranmer, M, Steel, D, Pickles, A, Martin, D, Gardiner, C, Fieldhouse, E, Elliot, M, Dale, A, Brown, M, (forthcoming) 'Microdata for Small Areas', *Royal Statistical Society Journal (A).*