A One Number Census: Proceedings of a Research Workshop

University of Leeds, May 12-13 1998

Edited by Stephen Ludi Simpson

CCSR Occasional Paper 15 (Discussion Chapters)

.

.

CCSR OCCASIONAL PAPER 15 (DISCUSSION CHAPTERS) A One Number Census: Proceedings of a Research Workshop

Edited by Stephen Ludi Simpson

Main contributions by: James Brown (University of Southampton) Lisa Buckner (Office for National Statistics) Ray Chambers (University of Southampton) Roma Chappell (Office for National Statistics) John Charlton (Office for National Statistics) Marie Cruddas (Office for National Statistics) Ian Diamond (University of Southampton) Tim Jones (Office for National Statistics) Pat Mann (Office for National Statistics) Liz Middleton (University of Manchester) Martha Farnsworth Riche Stephen Ludi Simpson (University of Manchester/Bradford Council) Andy Teague (Office for National Statistics)

A report to the Economic and Social Research Council, supported by ESRC Awards R000236963 and H507265031

This report contains the presentations made and the discussions held on the first day of the fourth ESRC/JISC Workshop Planning for the 2001 Census. It presents the views of expert census users on the need to make census output consistent with population estimates by accurately estimating census undercount, and on the current ONS plans in this regard for the 2001 Census.



CONTENTS

INTR Steph	ODUCTION en Ludi Simpson and Phil Rees	iii
Ackn List o List o	owledgements f speakers f participants	iv v vi
PART	1: CENSUS UNDERCOUNT. WHAT HAPPENED IN 1991? WHY DOES IT MATTER TO USERS TO USERS?	
1.	Characteristics Of Census Undercount; What Do We Know Already? Adjusting Census Output For Undercount Stephen Ludi Simpson and Liz Middleton	1
2	Simulation Of The Impact Of Non-Response On Census Applications Liz Middleton and Stephen Ludi Simpson	15
3	Invited discussant: Census Undercount. What Happened In 1991? Why Does It Matter To Users? Gillian Goddard	39
4	Discussion: Census Undercount. What Happened In 1991? Why Does It Matter To Users?	43
PART	2: OVERVIEW. DERIVING NATIONAL AND REGIONAL POPULATION ESTIMATES	
5	Overview Of 2001 One Number Census Proposals And Research Marie Cruddas	47
6	Design Of The Census Coverage Survey James Brown, Ray Chambers, Ian Diamond and Lisa Buckner	55
7	Dependence, Non-Response And Underenumeration Estimation Methods James Brown, Ray Chambers, Ian Diamond and Lisa Buckner	67
8	Demographic Analyses In Support Of A One Number Census Roma Chappell, John Charlton and Ian Diamond	79
9	Invited discussant: Overview. Deriving National And Regional Population Estimates Steve Turner	91

Page

10	Discussion: Overview. Deriving National And Regional Population Estimates	95	
PART 3: COMMUNICATION, POLITICS, AND THE CENSUS			
11	Communication, Politics, And The Census Martha Farnsworth Riche	99	
12	Discussion: Communication, Politics, And The Census	105	
PART	4: LOCAL POPULATION ESTIMATES: ADJUSTING THE CENSUS DATABASE		
13	Modelling Down To Small Areas James Brown, Ray Chambers, Ian Diamond and Lisa Buckner	107	
14	Small Area Estimation For A One Number Census - Weighting Vs. Imputation Ray Chambers, Marie Cruddas, Tim Jones	117	
15	Invited discussant: Local Population Estimates: Adjusting The Census Database Jenny Boag	127	
16	Discussion: Local Population Estimates. Adjusting The Census Database.	129	
PART 5: MANAGEMENT ISSUES			
17	Maximising Coverage In The Census Enumeration - An Operational View Pat Mann	133	
18	One Number Census And Census Coverage Survey Management Issues Andy Teague	137	
19	Invited Discussant: Management Issues Keith Dugmore	141	
20	Discussion: Management Issues	143	

INTRODUCTION

Stephen Ludi Simpson and Phil Rees

The Population Census is a very large exercise in data collection and processing. In 2001 some 25 million households will be contacted and asked to provide answers to 25 to 30 questions about 60 million persons living in the UK. Such a task is likely to cost £125-£150 millions. It is essential therefore that the Population Census is very carefully planned beforehand, and that the maximum information value is extracted from the data collected.

Output statistics from the Population Censuses are purchased by central government, local government, the health service, businesses, research organisations and the academic community. On the basis of this output, many billions of pounds are distributed to public services, while major public and private investments are located at specific sites. Social policy is shaped by the trends and inequalities revealed by the census.

The UK Census Offices consult with users to help plan the questions for the census and the outputs from it that are most appropriate to its applications. The completeness of the data collected from each small part of the UK is the distinctive value of the national census, which has achieved 98% coverage of the population in both 1981 and 1991. However, concerns that the differential nature of non-response affected the fundamental quality of the census led ONS to make the limitation and measurement of non-response a major priority in the planning of the 2001 Census.

This paper reports on presentations and discussion on the impact of census undercount in 1991 and ONS plans to reduce its impact in the 2001 Census. In Part 1 two papers discuss the nature of non-response to the 1991 census, and simulate its impact on resource distribution and estimates of regional migration. In Part 2, papers describe the plans to estimate the 2001 census non-response for each country and each District. In Part 3, experience from the USA highlights the importance of convincing politicians and public as well as technical users of data, that adjustments for non-response are appropriate. In Part 4 the papers are devoted to the means by which estimates for smaller areas may be made, and options to ensure that all census output for small areas consistently includes an allowance for non-response, which would be a major world-wide innovation. Part 5 is concerned with the reduction of non-response by good census fieldwork, and the management of the follow-up Census Coverage Survey.

Within each part the Workshop discussions are reported. This report of the presentations and the discussion is intended to allow others to understand the UK Census Offices' plans which will be put to Parliament in outline at the end of 1998, and to give the UK Census Offices feedback on users' approval and concerns with the plans and their impact on their work.

ACKNOWLEDGEMENTS

The programme for this workshop was organised by the research project *Quantifying the impact of census non-response on social policy and social research* supported under ESRC Award R000236963. The workshop itself became one of the series of *Workshops Planning for the 2001 Census - Determining Academic Community Needs and Strategy*, supported by ESRC and JISC under ESRC Award H507265031.

The UK Census Offices have been able to accept invitations to attend the Workshops and receive the inevitable mix of brickbats and bouquets. In particular those working on Census Plans at ONS in Titchfield and the University of Southampton's Social Statistics Department presented their work in a good spirit of constructive consultation. Papers in parts 2 and 4 are mainly taken from the *Census Consultation Paper - 2001 A One Number Census* published by the Government Statistical Service in April 1998. References from these papers are to be found at the end of chapter 14.

Several colleagues gave their time to chair the five sessions of the Workshop and thanks are due to Daniel Dorling (University of Bristol), John Hollis (London Research Centre), Andy Teague (ONS), Bruce Penhale (Oldham Council) and Tony Champion (University of Newcastle upon Tyne). Jan Howard, Paul Williamson, Eileen Howes, and Richard Arnold took notes of the lively discussion. They and the participants take credit for bringing forward the ideas within that discussion. The editor hopes there are no errors in the presentation of those ideas within this report, and has full responsibility for any such errors. The administrative arrangements for the Workshop were very efficiently executed by Christine Macdonald (University of Leeds), using the meeting rooms and accommodation facilities of Leeds University. This report of the workshop was ably brought together by Margaret Murray (University of Manchester) from a variety of scripts and annotations.

List of Speakers

James Brown

Social Statistics, University of Southampton, Highfield, Southampton SO17 1BJ Tel: 01703 593070, Fax: 01703 593846, Email: jjb1@soton.ac.uk

Lisa Buckner

ONS, Rm 4200W, Segensworth Road, Titchfield, Fareham, Hampshire PO15 5RR Tel: 01329 81 3507, Fax: 01329 81 3638, Email: lisa.buckner@ons.gov.uk

Ray Chambers

Social Statistics, University of Southampton, Highfield, Southampton SO17 1BJ Tel: 01703 594 311, Fax: 01703 593 846, Email: rc@alcd.soton.ac.uk

Roma Chappell

Population Estimates Unit, Room 2300, ONS, Segensworth Road, Titchfield, Fareham PO15 5RR. Tel: 01329 813 262, Fax: 01329 813 295, Email: roma.chappell@ons.gov.uk

John Charlton

ONS, 1 Drummond Gate, London SW1V 2QQ Tel: 0171 533 6239, Fax: 0170 533 6238, Email: john.charlton@ons.gov.uk

Marie Cruddas

ONS, Segensworth Road, Titchfield, Fareham PO15 5RR Tel: 01329 81512, Fax: 01329 813638, Email: marie.cruddas@ons.gov.uk

Ian Diamond

Social Statistics, University of Southampton, Highfield, Southampton SO17 1BJ Tel: 01703 592518, Fax: 01703 593 846, Email: idd@socsci.soton.ac.uk

Tim Jones

ONS, 1 Drummond Gate, London SW1V 2QQ Tel: 0171 533 6232, Fax: 0171 533 6238, Email: tim.jones@ons.gov.uk

Pat Mann

ONS, Segensworth Road, Titchfield, Fareham PO15 5RR Tel: 01329 81 3733, Fax: 01329 81 3532, Email: pat.mann@ons.gov.uk

Liz Middleton

CCSR, Faculty of Economics and Social Studies, University of Manchester, Dover Street, Manchester M13 9PL. Tel: 0161 275 4736, Fax: 0161 275 4722, Email: e.middleton@man.ac.uk

Martha Farnsworth Riche Email: mriche@erols.com

Stephen Ludi Simpson Bradford Council and CCSR, Faculty of Economics and Social Studies, University of Manchester, Dover Street, Manchester M13 9PL Tel: 01274 642 838, Fax: 01274 752 004, Email: ludi@man.ac.uk

Andy Teague ONS, Census, Titchfield, Fareham PO15 5RR Tel: 01329 81 3404, Fax: 01329 81 3407, Email: andy.teague@ons.gov.uk

List of Participants

Seraphim Alvanides School of Geography, University of Leeds, Leeds LS2 9JT

Felicity Andrew PPRU, Oxfordshire CC, County Hall, Oxford OX1 1ND

Richard Arnold Epidemiology & Public Health, Imperial College, Norfolk Place, London W2 1PG

Colin Bainbridge Policy Development, North Yorkshire CC, County Hall, Northallerton DL7 8AH

Greg Ball The Environment Department, Shropshire CC, The Shirehall, Abbey Foregate Shrewsbury SY2 6ND

Robert Barr School of Geography, University of Manchester, Oxford Road, Manchester M13 9PL

Robert Beatty Northern Ireland Statistics and Research, Arches Centre, 11-13 Bloomfield Avenue, BELFAST BT5 5HD

Amanda Bellringer GRO(S), Ladywell House, Ladywell Road, Edinburgh EH12 7TF

Jennifer Boag, Falkirk Council, Chief Executives Office, Municipal Building, Falkirk FK1 5RS

John Boyle Economic Affairs Division, Illiam Done House, 2 Circular Road, Douglas, Isle of Man, IM1 1 PQ

Paul Boyle Geography, University of Leeds, Leeds LS2 9JT

Chris Brunsdon Town & Country Planning, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU

Julian Calder ONS, D2/09, 1 Drummond Gate, London SW1V 2QQ

Jackie Carter Manchester Computing, University of Manchester, Oxford Road, Manchester M13 9PL

Tony Champion Geography, Daysh Building, Newcastle upon Tyne NE1 7RU

Martin Charlton Geography, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU

Jim Clark GRO(S), Room 1/1/1, Ladywell House, Ladywell Road, Edinburgh EH12 7TF Samantha Cockings Epidemiology & Public Health, Imperial College at St Mary's, Norfolk Place, London W2 1PG

Garnet Compton GRO(s), Ladywell House, Ladywell Road, Edinburgh EH12 7TF

Mike Coombes

NERRL, CURDS, University of Newcastle, Claremont Bridge, Newcastle upon Tyne NE1 7RU

Angela Dale CCSR, Faculty of Economics & Social Studies, University of Manchester, Oxford Road, Manchester M13 9PL

Hywel Davies Demographic & Social Studies, LRC, 81 Black Prince Road, London SE1 7SZ

Chris Denham Census Services Branch, ONS, Segensworth Road, Titchfield, Fareham PO15 5RR

Roger Denton Norwich CC, Chief Executives Dept, City Hall, Norwich NR2 3BE

Roger Dewhurst 49 Brownsvill Road, Heaton Moor, Stockport, Greater Manchester SK4 4PF

John Dixie ONS, Segensworth Road, Titchfield, Fareham PO15 5RR

Brian Dodgeon ONS Longitudinal Study Support Team, SSRU, City University, Northampton Square, London EC1V 0BH

Daniel Dorling Geography, University of Bristol, University Road, Bristol BS8 1SS

Keith Dugmore Demographic Decisions Ltd, 8 Hugh Street, London SW1V 1RP

Oliver Duke-Williams School of Geography, University of Leeds, Leeds LS2 9JT

Jason Dykes Geography, University of Leicester, Bennet Building, Leicester LE1 7RU

Aodan Edmonds Geography, Trinity College, Dublin, Ireland

Piers Elias Tees Valley Joint Strategy Unit, PO Box 199, Melrose House, Melrose Street Middlesborough TS1 2XF

Heather Eyre Geography, University of Leeds, Leeds LS2 9JT

Jan Finnigan

Tyne & Wear Research & Information, Room 146, Civic Centre, Newcastle upon Tyne NE1 8QN

Graham Foster Directorate of Planning Services, The Town Hall, Hornton Street, London W8 7HX

A Stewart Fotheringham North East Regional Research Lab, CURDS, University of Newcastle, Newcastle upon Tyne NE1 7RU

John Fox ONS, 1 Drummond Gate, London SW1V 2QQ

Forest Frankovitch NHS Executive North West, 930-932 Birchwood Boulevard, Birchwood, Warrington WA3 7QW

Jan Freeke Planning, Glasgow City Council, 231 George Street, Glasgow G1 1RX

Martin Frost Geography, Kings College London, The Strand, London WC2R 2LS

Linda Frost Planning Studies, Environment & Development, Manchester City Council, PO Box 463, Manchester M60 3NY

Chris Gardiner School of Urban and Regional Studies, Sheffield Hallam University, Sheffield S1 1WB

Isobel Gibson 41 Jubilee Close, Ledbury, Herefordshire, HR8 2XA

Gillian Goddard Room 443B, Department of Health, Skipton House, 80 London Road, London SE1 6LH

Robin Hall Planning and Estate Department, Chesterfield Borough Council, Town Hall, Chesterfield, S40 1LP

James Harris Manchester Computing, University of Manchester, Oxford Road, Manchester, M13 9PL

Tom Hennell NHS Executive North West, 930-932 Birchwood Boulevard, Birchwood, Warrington WA3 7QW

John Hollis London Research Centre, 81 Black Prince Road, London SE1 7SZ

Stuart Holroyd The Environment Department, Shropshire County Council, The Shirehall, Abbey Foregate, Shrewsbury SY2 6ND

Jan Howard Land Use Strategy, Milton Keynes Council, Civic Offices, PO Box 114, 1 Saxon Gate East, Milton Keynes MK9 3HQ

Eileen Howes London Research Centre, 81 Black Prince Road, London SE1 7SZ Norman Jamieson City of Edinburgh Council, City Development Department, 1 Cockburn Street, Edinburgh EH1 1ZH

Stephen Jones Kirklees Metropolitan Council, Devt. Unit, 2nd Floor, Civic Centre 111, Huddersfield HD1 2EY

Graham C Jones Room 4300W, ONS, Segensworth Road, Titchfield, Fareham PO15 5RR

Heather Joshi SSRU, City University, Northampton Square, London EC1V OHB

Mandy Kingston The NENE Centre for Research, NENE University College Northampton, Park Campus, Boughton Green Road, Northampton NN2 7AH

Catherine Knight Research and Intelligence Unit, Cheshire County Council, County Hall, Chester CH1 1SF

Neil Lander-Brinkley ONS, Census, Population & Health Group, Segensworth Road, Titchfield, Fareham PO16 7UU

Yong Lee

Tyne & Wear Research Information, Room 146, Civic Centre, Newcastle upon Tyne NE1 8QN

Rachel Leeser London Research Centre, 81 Black Prince Road, London SE1 7SZ

Rob Lewis London Research Centre, Parliament House, 81 Black Prince Road, London SE1 7SZ

David Lloyd Prescribing Support Unit, Brunswich Court, Bridge Street, Leeds LS2 7RJ

Kevin Lynch

Social Statistics Research Unit, The City University, Northampton Square, London EC1V OHB

Christine MacDonald School of Geography, University of Leeds, Leeds LS2 9JT

Kenneth MacKinnon Ivy Cottage, Ferintosh, The Black Isle, Ross-shire IV7 8HX

Nick MacReady Ordnance Survey, Romsey Road, Southampton SO16 4GU

David Martin Geography, University of Southampton, Highfield, Southampton SO17 1BJ

Peter Maxson Planning, Trans., & Economic Strategy, Warwickshire County Council, PO Box 43, Shire Hall, Warwick CV34 4SX

Mark McConaghy B4/11, ONS, Drummond Gate, Pimlico, London SW1V 2QQ Robert Moore Census Advisory Committee, Bennachie, Carmel Road, Holywell, Flintshire CH8 7DD

Roger V Morgan Room 328, The Town Hall, Hornton Street, London WB 7NX

Donald Morse UKBORDERS Service, Data Library, Computing Service, University of Edinburgh, George Square, Edinburgh EH8 9LJ

Barbara Noble DETR, Great Minster House, 76 Marsham Street, London SW1P 4DR

Andrew Osborn EDS, 4 Roundwood Avenue, Stockley Park, Uxbridge UB11 1B2

Matthew Ovington Department of Geography, Trinity College, Dublin, Ireland

David Owen Centre for Research in Ethnic Relations, University of Warwick, Coventry CV4 7AL

Ceri Peach School of Geography, Mansfield Road, Oxford OX1 3TB

Andrew Peloe School of Geography, University of Leeds, Leeds LS2 9JT

Bruce Penhale Chief Executives Department, Oldham MBC, PO Box 160, Oldham OL1 1UG

Simon Pitts Policy Development, Environmental Services, Yorkshire County Council, County Hall, Northallerton DL7 8AH

Simon Prinn Welfare Rights Unit, Ponsonby House, Edward Street, Stockport SK1 3PT

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

Christine Reid Nottingham City Council, Housing, Policy Section, 14 Hounds Gate, Nottingham, NG1 7BA

Sheila Ritchie Planning Studies, Environment & Development, Manchester City Council, PO Box 463, Manchester M60 3NY

David Rossiter OUCS, 13 Banbury Road, Oxford OX2 6NN

Krysia Rybaczuk Geography, Trinity College, Dublin, Ireland Joanna Schmidt University Computing Service, University of Leeds, Leeds LS2 9JT

Glen Scrivner Prescribing Support Unit, Brunswick Court, Bridge Street, Leeds LS2 7RJ

Andrew Shackleton

Corporate Development Unit, Kirklees M.C., Civic Centre 111, Huddersfield HD1 2EY

Jon Simonds C91 Project, County Hall, Powys County Council, Llandrindod Wells, Powys LD1 5LG

Jillian Smith B7/11, ONS, 1 Drummond Gate, London SW1V 2QQ

Terry B Smith Leeds City Planning Department, 44 The Headrow, Leeds LS1 8HR

Joanna Southworth Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS

Malcolm Spittle Policy Development Unit, Environmental Sciences, North Yorkshire County Council County Hall, Northallerton DL7 8AH

Frank Thomas GRO(S) Census, Ladywell House, Ladywell Road, Edinburgh EH12 7TF

Alistair Towers Data Library, Main Library, George Square, Edinburgh EH8 9LJ

Eddie Turnbull GRO(S), Ladywell House, Ladywell Road, Edinburgh EH12 7TF

Steve Turner

Tees Valley Joint Strategy Unit, PO Box 199, Melrose House, Melrose Street, Midlesborough TS1 2XF

Ian Turton School of Geography, University of Leeds, Leeds LS2 9JT

David Vallance Salford Software Ltd, Adelphi House, Adelphi Street, Salford M3 6EN

Nigel Walford School of Geography, Kingston Polytechnic, Penrhyn Road, Kingston Upon Thames Surrey, KT1 2EE

Judith Walton ONS, Room 2300, Segensworth Road, Titchfield, Fareham PO15 5RR

Margaret Ward The Data Archive, University of Essex, Wivenhoe Park, Colchester CO4 3SQ

Simon Whalley Flat 4, 199 Kirkstall Lane, Headingley, Leeds LS6 3EJ Dick Wiggins Sociology, City University, Northampton Square, London EC1V 0HB

Paul Williamson Geography, University of Liverpool, Roxby Building, Liverpool L69 3BX

Tom Wilson School of Geography, University of Leeds, Leeds, LS2 9JT

Bill Wood Chief Executives Office, Essex County Council, County Hall, Chelmsford, Essex CM1 1LX

CHAPTER 3 INVITED DISCUSSANT: CENSUS UNDERCOUNT. WHAT HAPPENED IN 1991? WHY DOES IT MATTER TO USERS?

Gillian Goddard

Introduction

My own experience of Census data and related issues goes back to the 1970s and 1980s when I had been involved in a variety of Census collecting and using activities. My use lapsed during the 1990s, leaving imprinted in my mind a belief in Census quality, what we have often referred to as the 'gold standard'. While I recognised it was not perfect, nor never could be - for the majority of practical purposes and uses it was as good as it could be and its imperfections were minor compared with those of other sources.

As these two papers demonstrate, 1991 was not thought to be up to the same standard. So this workshop of Census experts, users and collectors are considering the improvements for 2001.

1. I agree with the authors of these papers that much of the work has to be done by ONS at source using individual records. The results of making plausible or very plausible adjustments show what can be done with considerable effort. Who wants to regularly invoke the Iterative Proportional Fitting (IPF) procedure? Not I.

2. So ONS must do whatever is needed to convince the user that the data are meaningful, accurate for the uses required and unbiased. Public perception must be changed. Users now are more difficult to please, better educated and aware. They will need convincing that the Census data are fit for purpose. The government needs accurate statistics about the population on which to make a whole range of decisions, allocate resources and evaluate policy. Increasingly there is an expectation that we can count and measure a range of detailed characteristics down to small relatively small areas. In many cases, the 1991 Census data did fill this aim, but some questions remain.

3. From a Department of Health perspective I am aware of concerns over how the characteristics of those that were missed may significantly affect the interpretation. I have not tested the effects in resource allocation by DH but am reassured by the SSA analysis that the effects are likely to be small. Population is not the only component and it is used in such a way that such variations may well not affect the eventual outcomes.

4. But this may not be true for all data uses. The example of the missing elderly and the effects on mortality rate is relevant because it shows how underenumeration will lead to higher mortality rates. Deaths are well recorded so if populations are too low the rates will be too high. These in turn will feed into population estimates and projections which may well be

biased at older ages. The effects will be more noticeable by area if there is differential response by area, which seems very likely.

5. Policy makers in the Department of Health have already asked me to explain how the 1991 Census based data allow for underenumeration and non-response. I will briefly run through some of these points using the material from ONS, but supplementing them to raise questions relevant to 2001 issues.

6. Post 1991 adjustments which are affected by underenumeration and non-response:

- students and boarding school pupils who were recorded as at their home address in the Census had to be adjusted to their term time address for population estimates; Question how well were they counted? Can we ever ensure we count them where we want them (at term addresses). Will changes proposed for 2001 improve matters? Following later discussions I became more concerned that enumerating them through the Census Coverage Survey (CCS) was fraught with difficulty as the timing of the CCS 4 to 6 weeks after Census day is likely to coincide with them not being at their term addresses.
- visitors with no, or no identifiable, usual address these were people recorded at an address other than their usual residence on Census night; Can we be sure we get them at their usual address under the 2001 proposals? Will these be missed in both Census and CCS?
- adjustments for under-enumeration made from the Census validation survey. Will the 2001 Census coverage survey (CCS) succeed where the CVS failed. Will we ever get the really hard to count?
- enhancement of the count of Armed forces (and dependents of foreign forces) using MOD data, and more recently using Defence Analytical Services Agency (DASA) data for numbers of Armed forces; maybe figures will improve for 2001 - but it seems difficult to get data relating to where the Armed Forces live, as opposed to where they are based.
- modification of the count of elderly residents (aged 80 and over) using DSS pensioner data; - can we get better geographical data in 2001 to show where the missing elderly are? I am concerned about missing some of the single person households, particularly the elderly who may well be visiting elsewhere or in hospital around Census day. I remain hopeful that the Department of Health will have some better data available to help validate aggregate records.

• enhancement of the national count of people aged 1-44 to allow for additional underenumeration. This was found to be necessary by comparing the results from the previous census of those 10 years younger. This revealed a shortfall of people aged 1 to 44 (particularly males aged 20-29) that could not be explained. What will we do for 2001? How will we get them to the right part of the country?

7. I considered the importance of the Census migration question (address one year ago). It is used in population projections to help distribute subsequent movers and migrants for whom there is no more detailed source. Underenumeration results in perpetuating some of the deficiencies as the young mobile - missed - group cannot accurately contribute to this. It is also likely to be deficient for the people who are asylum seekers or visitor switchers. This latter group comprises those who intend to stay for less than one year, but later apply to stay for one year or longer. They can include those marrying and students. They too are more likely to be in the 'missed' group. There is a need to get good information on them and their movement patterns to use in the future. The confirms views put later that the CCS must ask for address one year ago in order to allow more accurate data to be collected in 2001.

8. Ludi's paper states 'any remaining non-response is assumed to be random' which then allows the appropriate calculations to be done. But of course, it is not random. The very hard to count may well not display the characteristics of those we think to be like them but counted. Their health, migration and a whole variety of other characteristics could well be different.

9. We have assumed that the institutional population has been fully counted, yet the armed forces were not. That is a worrying discovery and should cast doubt on the quality of some other institutional data. Later in discussion the quality of the prison population was questioned. I wonder why the demographic characteristics of this group, if nothing else, cannot be accurately recorded.

10. It is good to have all these thoughts in mind as we move on to a fuller discussion and the rest of the workshop. My only feeling is that we may have to do more than work with the best that a Census collection can produce these days and continue to milk any other sources available. Some of the hard to count will want to receive grants and forms of welfare which generate some alternative administrative sources and sadly the poor and elderly will be more likely to encounter the health service which can demonstrate a need to know about some of the demographic characteristics of those it assists.

11. I also want to make users aware of the potential of the National Strategic Tracing Service of patient records. I have a short paper describing it. I agree it is not going to be in place beforehand in order to assist at this stage but might well provide useful aggregate data at a later stage to help ONS ensure that the one number is the right one.

CHAPTER 4 DISCUSSION: CENSUS UNDERCOUNT. WHAT HAPPENED IN 1991? WHY DOES IT MATTER TO USERS?

Extreme adjustments

Phil Rees and others questioned the extreme adjustments made in the simulation. Ludi noted that the impact all the census output examined was not in fact much greater when simulating extreme levels of non-response than for the medium levels, nor in the 'combined' model where all three of employment status, tenure and ethnic group affected non-response. Summary tables in the second paper had used the medium combined levels of non-response, because this reflected whichever of the three social factors was relevant, without giving extreme and implausible results.

The need to spend resources on IPF to estimate census tables for each ward of the country before adjusting them, would be removed if these had been commissioned from the census offices, but the cost was currently prohibitive, several thousand pounds for each different table, with considerable delays between order and supply.

How bad is the impact of non-response?

It was suggested that the decision to be made was whether the UK's coverage was as good as other countries', or 'other countries were as bad as the UK'. Keith Dugmore felt that the attention given to non-response, concentrating on the worst effects, was out of all proportion to the problem for most users, and that this attention detracts from the whole Census.

Generally the results from the project so far were thought to show that the impact on spending assessment were minimal, but the major part of the SSA had not yet been simulated.

Why reduce non-response?

Roger Morgan suggested those that did not return a form did not deserve services. On the other hand, the impact was not so much on them as on others in the same area who had responded.

Good enumeration in the first place

The ONS aimed to make the form easier for people to complete, to count people only once, and to simplify the job of the enumerators (Andy Teague). Keith Dugmore pressed for better communication about the nature and purpose of the census. Isobel Gibson felt that differential non-response could also be reduced by reducing variance between enumerators, through training.

Residents temporarily abroad

Pose particular problems for coverage of both the Census and its coverage survey (Tom Hennell)

Was 1991 special? Will 2001 be special too?

Paul Boyle suggested that it may have been, due to the campaigns against the recentlyintroduced poll tax. However, the characteristics estimated for the census non-response are generally found in other countries an in Britain in other settings and at other times when no such campaigns were held.

Danny Dorling pressed that the attitude to government enquiries in 2001 should be kept in focus. For example recent low turnouts in London elections should be a warning, because voting turnout is likely to be correlated with census non-response. Similarly, the first students with debts from student loans will be graduating in 2001.

Metadata about the adjustments

John Fox said that public perception was affected by the debate, and that we need to get back to accepting 2001 census output as the 'Gold standard'. He said that there was a need for adjustments to be done centrally but quickly. The methodology must be agreed in advance and we must think about the characteristics of adjustment methods.

He questioned what information or metadata users would want to enable them to use the Census data with more confidence. What type of confidence intervals were needed? Several contributors supported his point that users needed ways of coping with uncertainty in the data they used, even when it is adjusted to be unbiased.

Adjust only the age-sex structure of the population?

To the suggestion that it may not be necessary to adjust characteristics of the population, Sheila Ritchie stated that grants are allocated to particular groups on the basis of the Census data, that these adjustments could not be made locally - although politicians demanded that it be done in order to exploit weaknesses in the data.

Adjusting small cells

Chris Denham felt that adjustments to most small area data would be estimated as small fractions of people which would either be rounded out of existence, or imputed as whole people in which case would have a lot of inaccuracy attached.

CHAPTER 9 INVITED DISCUSSANT: OVERVIEW. DERIVING NATIONAL AND REGIONAL POPULATION ESTIMATES

Steve Turner

1. Introduction

- 1.1 My aim in this part of the session is to review the previous very interesting papers and give a personal slant on the issues they raise. It will tend to be from a local authority perspective. This is a world where people want answers to problems. They do not care if it comes from the Census or Pandora's box, as long as it helps them understand. Nor do they want ifs and buts, may be this or may be that. They want something that is simple - preferably one number.
- 1.2 In the spirit of retaining simplicity, there will be no mathematics in this paper. I can generally understand the type of mathematics used in previous papers but I am leaving any constructive criticism on that front to others. I deal with the positive side of the work so far, then cover a number of concerns followed by some final comments.

2. Why The ONC Concept Is Worthwhile

- 2.1 I think the greatest strength of the One Number Census process is represented by the fact that we are discussing it now, in May 1998. Not only that, we also have a very substantial body of work already in place or under way as demonstrated by the papers under consideration.
- 2.2 Throughout my involvement with the process, the commitment by ONS has been very apparent. It (or rather, its forerunner OPCS) was stung by the loss of confidence by the user community in the population estimates after 1991. A firefighting exercise was necessary then and ONS clearly wants and needs to 'get it right' next time round after 2001. Commitment is evidenced by the ONC process having high level involvement within ONS, having a substantial budget devoted to investigating it and that overall census planning has the ONC principle as an integral part of the process it is not a 'bolt on' element.
- 2.3 It also has to be said that the firefighter turned fire prevention officer Ian Diamond with his colleagues from Southampton University in tandem with the researchers at ONS, have made very significant progress. Their application and ingenuity have, I believe, produced an excellent foundation for what has to follow in the ONC process.

- 2.4 On that score, I thought I would pause to contemplate what we want from a census in very general terms. I suggest it should count the number of people in each area and give a range of socio-economic information. Well, it certainly gives large quantities of the latter which area has most unemployment, which has the lowest car ownership but it does not actually give a proper estimate of population size, because of underenumeration.
- 2.5 I believe the ordinary user expects the census to give an accurate estimate of population size. Indeed, the ordinary user typically thinks he or she has it when using census data. In a sense, previous censuses have been fraudulent because they have not actually fulfilled this expectation. Of course, there has been a process whereby population estimates have been produced by ONS. Amongst other uses, they figure in the Standard Spending Assessment and thus contribute to the calculations which result in the distribution of billions of pounds in grants for local authorities.
- 2.6 A successful ONC process would give much improved population estimates at local authority level which would be well worthwhile given the sums involved in grant allocations. If a full ONC is produced then I think it will help instil more confidence in users. The data they would then be using would represent the actual number of people in the area. The more experienced census user would not have to give an embarrassed cough and attempt to explain why some people are missing or, more likely, sweep the issue under the carpet.

3. Why I Am Still Worried

- 3.1 My worries are not really about what has been achieved so far in the process. They largely amount to a number of concerns on what still has to be done. There are new processes to be tested and more mathematics to be devised.
- 3.2 As I understand it, the Census Coverage Survey is planned to be the key to producing better population estimates. A sample of areas is to be completely enumerated for certain basic variables. People enumerated in the CCS are to be matched with those enumerated in the Census in the same areas. Combining the two sets of data, the mathematics will be designed to cope with people missed by both and the results should be improved population estimates nationally and at local authority district level. Moreover, it is hoped to extend this process to alter all census counts to give the 'One Number'.
- 3.3 Many of the following concerns deal with minimising potential errors for very small percentages of the population. However, given that our concern is for the 2% or so underenumeration, all are likely to be relatively significant. Indeed, they are likely to be of most concern in the very areas where the highest underenumeration occurs.

- 3.4 Here are the issues which worry me:
 - The CCS itself has only been tested on a small area so far.
 - The process of identifying the post codes for the CCS in the field has to be exact. The maps have to be of high quality to enable unambiguous identification of the areas to be enumerated and the enumerators must use them accurately.
 - It is vital to get a high CCS response rate given that (unlike the Census itself) it is voluntary. Response will probably be lowest in the very areas where the Census itself is lacking.
 - The CCS is being conducted around 4 weeks after the census and around 0.8% of the country's population are likely to have moved in that time. How will they be accounted for? Will this lead to overcounting in the resulting population estimate if immigrants in the CCS are counted.
 - The process of matching individuals and households between the Census and the CCS is critical. In particular, if an individual enumerated in both is not matched, then that person is counted twice and the census undercount potentially becomes an overcount. Clearly, this has to be minimised but there will be individuals falling into this category. The process and the mathematics has to allow for this.
 - I find the work on the demographic estimates to be valuable. However, it is distressing but I accept with resignation that the International Passenger Survey has had to be used to estimate international migration. I would hope there would be a longer term aim to enhance the IPS significantly or replace it with a better source of migration data.
 - What will actually be the process when the national census-based estimate is compared with the national demographic estimate. Which will be chosen as the 'winner' or will a third compromise number nose ahead?
- 3.5 I find it a shame that a third source of data of sufficient quality was not available. This would have been the key to the Triple System Estimator and potentially to enhance population estimates. The problem was that none of the administrative lists examined were of sufficient quality. I think that it is time for ONS to commence a programme which builds on the research they have already undertaken for ONC. This programme would examine all possible sets of administrative records with a view to finding those that could usefully be enhanced without prejudicing their designed purpose and used in a 2011 TSE. Potentially, such data sources could also be used in the inter-censal population estimates from 2002 to 2010.

- 3.6 A further major concern is the delay there will be in sending the census results to us, the users, as a result of including the ONC process. On the one hand, not all users of data will benefit from ONC. Indeed, some of my typical census users are becoming a little sceptical of my protestations that seven year old information still has relevance and meaning for their situations - they will not want to wait. I accept that some delay is worthwhile. However, I had not decided how long would be acceptable on first contemplating the issue two years ago and I am no closer to an answer now.
- 3.7 The final concern is that, for any of several reasons, the ONC process stops after the national and local authority district population estimates have been produced, and the census tables are produced unadjusted in 'traditional' form. If this were to be the case, many of the benefits of the ONC concept would be lost. In fact, it would not be an ONC One Number Census it would actually be an IPEC Improved Population Estimates Census since the population estimate and the census population would not correspond.

4. Final Comments

- 4.1 Overall, I find the progress so far to be encouraging and I am looking forward, in anticipation, to see how the remaining issues are tackled. In terms of presentation, I think the full exposure of the mathematics is essential. However, there may well be scope for satisfying a significant group of census users by having a version where the meaning of the mathematics is explained without their having to understand it. This might apply both to the work so far and to future work.
- 4.2 Overall, I find I have to reflect that this is a lot of effort to sort out a mere 2% or whatever of the population, in 2001. Nevertheless, I conclude that it is a necessary part of a concern for quality. I think that ONS and the rest of us in the 'concerned census community' have a responsibility to aim for that quality and to make it simple for the ordinary census user.

CHAPTER 10 DISCUSSION: OVERVIEW. DERIVING NATIONAL AND REGIONAL POPULATION ESTIMATES

The need to achieve a high degree of matching between census and CCS. Failure to match census and coverage survey, causing an over-estimate of the true population, particularly those who had moved during the 3-6 week interval, was highlighted as a danger by several participants. James Brown noted that the CCS will ask who were usual residents on Census night; potential problems in this question will need to be minimised through fieldwork training.

If the corrected census doesn't agree with demographic estimates - who decides then? Danny Dorling, Frank Thomas and others questioned what the means of agreement could be if the national corrected estimate is outside the confidence interval around the national demographic estimate. The assumption appeared to be that as in 1991 the Census would be assumed to be wrong if it were outside the IPS confidence limits. Is this so?

In response, Ian Diamond suggested the reconciliation would be done on a case by case basis, searching for errors in either data where there were differences.

A plain English version of the difficult maths. Several participants advised a version of the proposals in which the maths is explained without having to be followed mathematically.

Triple estimation with administrative registers. There was support for Steve Turner's suggestion of a programme of research and improvement of administrative registers so that they can be used in intercensal estimates 2001-2011 and in validation of the 2011 census, but general understanding that their quality was not sufficient for use in triple system estimation in the 2001 census.

Tom Hennell asked for clarification of why the NHS strategic tracing service would not provide, by 2001, an accurate administrative record suitable for triple estimation: was there anything that could make it better in time for use the census? Ian Diamond responded that the delay before people re-register with a GP after moving to their new address, means that the patient records could never be matched up with the census person-for-person, even if duplicate records were eliminated by improvements to the register. It may still be used however in the aggregate, to compare totals and age structures for areas.

Barbara Noble, a 'fire-fighter' in 1992, found then that none of the administrative sources was good enough. Schools data was and still is inadequate, and while DSS and NHSCR data may have improved somewhat FHSA data was not yet good enough. Barbara also put in a plea for a quality survey in addition to the coverage check, because information on quality would be essential for some questions.

Terminology for the one number census outputs. Phil Rees asked for clarification of the word 'estimate'. Robert Moore asked how government would counter the journalist who says "These figures have been cooked up. There was a million missing in 1991 and we are supposed to believe there is none missing in 2001. They are not counts". Stephen Ludi Simpson suggested the ONC estimates be distinguished from further work on population estimates by government, by calling them "Census population counts". James Brown suggested that these were communication issues which would be addressed as the methodology is fixed and 2001 approaches.

Roger Morgan was concerned about completely synthetic returns, and therefore imaginary people and whole households, and how they were to be allocated. How could he be reassured? (Chapter 14 covers Imputation vs Weighting).

Phil Rees further asked if users wanted a census day 2001 estimate or a June 30 2001 midyear estimate, to which there was no consensus.

Management and evaluation of the CCS. Keith Dugmore asked whether the Coverage Survey would be carried out in a different way to the Census itself, or if it would use the same mechanism. He was also concerned about the Coverage Survey being voluntary and the effect that might have on response rates. Why not massively target difficult areas, otherwise it will be difficult to get the 80% response that the presentation suggested is required. James Brown confirmed that field-workers will be put where they are most needed, requiring a flexible approach to fieldwork management.

Robert Beatty emphasised that the Coverage Survey was critical and must identify people not covered in the Census itself. How can you know whether the CCS has worked? James Brown: This will be difficult but nonetheless crucial to assess. I see it as coming from an examination of reports on the field implementation and refusals of the CCS with a comparison to census data and field information from the census itself. Assessing the performance of the CCS will also come from aggregate checks with demographic sources and other lists as mentioned earlier.

Confidence intervals in consultation and in output. Ian Diamond and James Brown displayed the implications of a relative standard error (RSE in the papers) of 2.5% for the design variable:

For a $\frac{1}{2}$ million population the 95% confidence interval for the overall total will be about $\pm 5,000 (\pm 1\%)$.

For a one million population the 95% confidence interval for the overall total will be about $\pm 7,000 (\pm 0.72\%)$.

For a five million population the 95% confidence interval for the overall total will be about $\pm 16,000 \ (\pm 0.32\%)$.

For the overall national population the 95% confidence interval will be about $\pm 50,000$ ($\pm 0.1\%$).

The regression models that can inform the calculation of confidence intervals are specific by age, sex, hard to count index, and separately estimated for each group of local authority districts.

Stewart Fotheringham asked whether people in local government wanted confidence intervals attached to Census data.

Others asked whether confidence intervals would be given for local or national, urban or rural, regional. Would different regression relationships be estimated from the CCS and Census in different areas?

Eileen Howes said that while some local authority users only wanted the Census results, those involved in local authority research and resource allocation needed the information confidence intervals would provide, even if they did not then publish a confidence interval with every count. This was especially important in allocation of resources and looking at the robustness of the data that allocations might be based upon.

Stephen Ludi Simpson said that expected confidence intervals were important in designing the One Number Census, and in approving the design now, and in evaluating the success of the One Number Census. He thought that at the output stage local authority users did not want confidence intervals around every figure but a straightforward way of calculating them approximately.

Roger Dewhurst supported the use of confidence intervals in a planning context.

Stephen Ludi Simpson asked for confidence intervals for age groups. He suggested that James Brown's boxplots be extended with simulated true and estimated counts by age for a typical sized-district, and be produced with numbers of people as well as with percentages. He thought this would help users understand just what error was being suggested as acceptable. James Brown felt this was a good way of communicating errors and agreed would be used when simulations extended to individual local authority districts as well as groups of local authority districts.

ONC adjustments may lead to a reduction in counts for some small populations though we will never know which ones. It was suggested that simulations could look for this effect in order to adjust the design in order to describe it and to avoid it as much as possible. James Brown reported that the ONC team were investigating gross errors and looking at how to avoid them through outlier checking and methods to deal with outliers. This is where comparison aggregate data from other sources will be potentially very useful.

CCS design. Using age structure alone to cluster the EDs within strata would not ensure homogeneous groups with respect to missingness. The ONC team would cluster on the hard to count index within strata, which would increase sample in the 'hardest to count' groups, though there is the issue of change since 1991 (see comment on the hard-to-count index below).

Households and household characteristics. Felicity Andrew and Rachel Leeser both raised the problems of household membership, and suggested that this should be a priority.

Rachel Leeser commented that the Census is about collecting information on how people live, and was not just a population count, so she and other users were particularly concerned about the proposed new household definition. The ONC would work with the household definition provided by ONS.

Postcode accuracy. David Martin raised the issue of the accuracy of postcodes and addresses within them. He felt that some addresses would still in 2001 be wrongly allocated, and that the Post Office might still make the same mistakes with new buildings as they currently do.

Dependence of CCS and Census. As a consequence of postcode (in) accuracy, David Martin pointed out that dodgy addresses not found by the enumerator will be the same addresses that the CCS will find with difficulty.

Stephen Ludi Simpson asked whether dependence of CCS and Census could be estimated, from a suitably designed CCS itself, in order to improve the estimates.

Random modification. Isobel Gibson and others did not want the data to be randomly modified after all the ONC adjustments.

Ian Diamond responded that, personally, he felt that confidentiality would be maintained without random data modification.

A record of adjustments. Ian Turton, Bruce Penhale and others were concerned that the true counts before the ONC exercise would still be available in addition to the adjusted data. James Brown responded: if we give you both counts will it be a ONC?

An up-to-date hard to count index. Bruce Penhale felt that conditions in some areas will have changed significantly since 1991: how would this affect the results? James Brown: By not weighting the design by the 1991 hard to count index the design will be robust to changes but not necessarily the most efficient if there has been no change. For estimation we can post-stratify by the new hard to count index defined by unadjusted 2001 Census counts which will reduce the chance of outliers in the regression models.

CHAPTER 12 DISCUSSION: COMMUNICATION, POLITICS, AND THE CENSUS

Martha Farnsworth Riche's was an evening talk without a discussant. She answered questions as follows

O: What is the response to the income question?

A: Martha Farnsworth Riche: That is the worst - we do more imputation on this than on any other question. People are more willing to tell us about their sex life than about their income. But income is required to implement legislation.

Q: Less than two years to go and there is still such uncertainty - is there any chance that the census won't happen in April 2000?

A: Martha Farnsworth Riche: There have been two recent tests, one with sampling and one without. The Autumn elections for Congress may determine whether we do it the old way or the new way. But the old way is not really possible - it would be a mess.

Q: At least in the USA you are legally required there to have a census, is that so?

A: Martha Farnsworth Riche: Well, there is one census - 1920 - which was taken but the results were not used. It was a time after immigration and industrialisation, and Congress didn't like the threat to the dominance of rural areas in favour of urban, foreign, wine-drinkers. This is looking a little the same, there's a lot of new people that are different to those in power.

Q: Are there any friends in the Senate? The Senate elections procedures are not dependent on Census results.

A: Martha Farnsworth Riche: There are five senators who are friends of statistics. One of them has said that it's almost impossible to get through a law that includes in it a budget for statistics to monitor the success of the law. This is partly because many senators are ex-Mayors, who want to reward their friends, and they know that the use of statistics will take away their power to do so.

CHAPTER 15 INVITED DISCUSSANT: LOCAL POPULATION ESTIMATES. ADJUSTING THE CENSUS DATABASE

Jenny Boag

There were four general points about the papers:

1. I am very impressed by the work which has gone into the One Number Census Project (ONC) and by the enthusiasm of the team doing the work. The methodology is continually developing, which makes commenting on it a little difficult, as comments seem to be overtaken by events! It is exciting to be at the leading edge of research. Most importantly, it looks like it is going to work.

2. This section of the ONC project, getting the data down to small area level, is the most important from a local authority point of view. If this part fails, then the whole project will have failed. The statistics at the local authority area are important but if that is all we get we are not much further forward than we were in 1991 when ONS/GRO(S) recalculated the mid year estimates of population. But the problems exist at a small area level. In my own council, the Census total population is just over 2,000 lower than the mid year estimate. This presents problems. Modelling down to small areas is essential.

3. We also need figures which add up. One of the problems with the 1991 Census is the Barnardisation which leads to data from different tables giving different figures. I checked the total number of households by adding about one third of the Output Areas in Falkirk and got figures for the total number of households which differed by over 200, depending on which tables I was using. This is not acceptable. We need to take the opportunity of the ONC to use the combination of the time delay between the date of the Census and the date of release of the data and the adjustment of the data through the ONC, to get rid of any further modification of the data at small area level.

4. We need to have an "idiots guide" to the methodology. The papers presented are very interesting but the maths and stats they contain are well beyond the ability of most Census users. In order to be accepted the ONC methodology needs to be understood. I appreciate that the methodology is still developing so it is probably too early to produce a simplified version. I also accept that the maths and stats detail is very important so that those who can understand it, can check its validity, but they will be very much a minority of users.

To go on to the main points of the papers; many of my concerns are acknowledged by the development team, but they will have to be addressed.

There doesn't seem to be a very clear answer on how to deal with the assumption that noone will be missed by the Census and the CCS. This seems to be part of the dual system

CHAPTER 15 INVITED DISCUSSANT: LOCAL POPULATION ESTIMATES. ADJUSTING THE CENSUS DATABASE

Jenny Boag

There were four general points about the papers:

1. I am very impressed by the work which has gone into the One Number Census Project (ONC) and by the enthusiasm of the team doing the work. The methodology is continually developing, which makes commenting on it a little difficult, as comments seem to be overtaken by events! It is exciting to be at the leading edge of research. Most importantly, it looks like it is going to work.

2. This section of the ONC project, getting the data down to small area level, is the most important from a local authority point of view. If this part fails, then the whole project will have failed. The statistics at the local authority area are important but if that is all we get we are not much further forward than we were in 1991 when ONS/GRO(S) recalculated the mid year estimates of population. But the problems exist at a small area level. In my own council, the Census total population is just over 2,000 lower than the mid year estimate. This presents problems. Modelling down to small areas is essential.

3. We also need figures which add up. One of the problems with the 1991 Census is the Barnardisation which leads to data from different tables giving different figures. I checked the total number of households by adding about one third of the Output Areas in Falkirk and got figures for the total number of households which differed by over 200, depending on which tables I was using. This is not acceptable. We need to take the opportunity of the ONC to use the combination of the time delay between the date of the Census and the date of release of the data and the adjustment of the data through the ONC, to get rid of any further modification of the data at small area level.

4. We need to have an "idiots guide" to the methodology. The papers presented are very interesting but the maths and stats they contain are well beyond the ability of most Census users. In order to be accepted the ONC methodology needs to be understood. I appreciate that the methodology is still developing so it is probably too early to produce a simplified version. I also accept that the maths and stats detail is very important so that those who can understand it, can check its validity, but they will be very much a minority of users.

To go on to the main points of the papers; many of my concerns are acknowledged by the development team, but they will have to be addressed.

There doesn't seem to be a very clear answer on how to deal with the assumption that noone will be missed by the Census and the CCS. This seems to be part of the dual system estimator approach, but the assumption is still being made as part of the developing methodology.

There is an issue about how to deal with smaller local authorities in the small area figures. There are three authorities in Scotland with a population of around 20,000, and there are also small authorities in Wales and in England. I am not convinced that the CCS methodology has addressed this issue properly yet. How these authorities will build up to pseudo-counties is a related issue.

The second paper dealt with the discussion of whether weighting or imputation was the best method of deriving small area data. Which ever method is chosen, some way of imputing characteristics not collected in the CCS needs to be developed. I think that, despite its problems, I would favour imputation.

We will need data which produces internally consistent individual and household level tabulations.

There seems to me to be a problem with weighting at a small area level. The very small numbers involved will mean that applying the weights will not necessarily lead to the addition of whole people with particular characteristics in tables as they will not reach the 0.5 threshold. I am also concerned as to how rounding will work at a small area level. Will tables add up consistently? Remember that Scottish Output Areas have an average of only 55 households or 120 people.

I would suggest that there are ways round the problem of imputing. I think there might be a workable alternative if the exercise were to start by calculating the number of households of different types (rather than sizes) and the number of individuals with different characteristics to be imputed; and imputed the households first, then subtracted the individuals imputed into these households, leaving the remainder of the individuals to be imputed into counted households. These could then be imputed on the basis of the characteristics of the households they should live in, by searching the database to find a matching household type. This would not necessarily have to be done at the smallest possible geographical level, but could be done by grouping postcodes by, for example, their hard to count index within a local authority.

Use could also be made of some of the information collected at enumeration. There should be lists of all households where a Census form was delivered and where one was not returned but the enumerator believed a household was resident. This could be used to impute missing households to an address with a known missing household, leaving any additional missing households calculated through the CCS to be imputed at random.

CHAPTER 16 DISCUSSION: LOCAL POPULATION ESTIMATES. ADJUSTING THE CENSUS DATABASE.

Imputation. Most participants who joined the discussion favoured imputation over weighting; in particular, as Danny Dorling pointed out, it seemed as though it was the best way of ensuring consistency between census outputs. In addition to the discussant's suggestions for imputation into households, use of enumerator reports on absent households as in 1991 was mentioned as essential information that so far did not figure in the ONC plans. The team responded that examination of the 1991 absent household imputation method will be part of the work into developing an imputation methodology for the ONC. It is has not yet been decided what information will be collected from absent households, but we would expect to make use of all additional information in carrying out imputation.

Imputation of small populations. Ken McKinnon was concerned that in 1991 imputation missed small populations - based on work with the 1991 Census Gallic population. Is this a bias in the method? Ian Diamond agreed that a careful imputation strategy was necessary for small groups - mentioning Somalis as a group whom others had claimed were underestimated in London.

Ian also felt that small LAs definitely posed a challenge to the estimation procedure.

Will every person be counted? The paper appears to assume that either the Census or the CCS will capture everyone in every household. Tom Hennell and others felt this was unjustifiable and must be relaxed in justification of future proposals. Ian Diamond responded that this assumption was relaxed in estimating design group totals, to which the local estimates are constrained.

Postcodes. Chris Denham was concerned that the Royal mails' postcode system is not yet up to the job required of it for the ONC, and the authors agreed that the quality of the postcode geography will affect the quality of the ONC results.

Weighting. Phil Rees commented that it is good to see use of statistical tools to increase completeness and validity of data, rather than preventing disclosures as has been more usual. The problem with use of Iterative Proportional Fitting (IPF) in weighting is that it does not give whole people. He offered his and Oliver Duke William's work connected to data disclosure, which ensures integer results from IPF for three dimensions.

Households. Ian Diamond re-iterated that synthetic households could not be used to make up all the missed people because some missed people were from existing households.

Jan Freeke stressed that comparability over time is important, for household types, and he felt that ONC proposals may endanger this.

Postcode population. Frank Thomas questioned the purpose of apparently estimating the total postcode population twice. James Brown: In a way we are for the sampled postcodes but for ONS purposes the information we get each time is very different. The second time round we are estimating a whole range of characteristics as well as the household structure where as the first time we are purely interested in population counts by age and sex.

More wanted on gross error. The authors mentioned a probability of 'gross error', i.e. error worse than the census, of 4/1000. Stephen Ludi Simpson felt that a small gross error such as this would convince users of the ONC claims. For what size population was this? For what size would this probability be only ½? Did the 'census' include imputed absent households?

James Brown responded: This came from the simulations for the individual age-sex groups in an LAD of size ¹/₂ million. As this is from the simulation, the results for smaller population subgroups will come as the work on small area estimation is tested through further more complex simulations.

John Dixie commented that there will always remain a degree of error even with a completely successful ONC, for example in the quality of responses. Discussion of other errors should be in this context, and will reduce the accuracy of imputation.

Differential uncertainty. Ian Diamond said in discussion "Confidence intervals will depend on the characteristics of the population you are looking at", while Stephen Ludi Simpson stressed that one of the primary aims of the ONC was to avoid differential uncertainty.

Simplify methods where possible: socio-spatial smoothing and mixed effects models. Stephen Ludi Simpson questioned the need for socio-spatial smoothing when predicting postcode undercount? Whatever this could achieve can equally well be achieved by including those same social variables as predictors in the regression model. James Brown responded that usefulness of the socio-spatial smoothing and how we might implement it is being investigated and it is by no means a certainty that this extra level of complexity will be worthwhile.

Stephen Ludi Simpson also asked if three would be any significantly different result using fixed effects for postcodes when estimating the regressions? If not, why introduce them? Wouldn't it be best to keep it simple? The team responded that they are examining the value added of random effects models over simpler fixed effects models and again we will be looking for significant gains in predictability from including the extra complexity.

Plain English justification. Paul Williamson, having talked to various participants after this session, felt that confusion reigned, some believing that weighting had been claimed to be most accurate, and others that imputation had the same claim. The model used by DETR for household projections which are also technically sophisticated, might be used: fulsome verbal descriptions of the methodology are included in appendices within the published report, but the gruesome detail of mathematics is saved for separate technical reports that are available to the 'junkies'.

CHAPTER 19 INVITED DISCUSSANT: MANAGEMENT ISSUES

Keith Dugmore

I'd like to thank Pat Mann and Andy Teague for their very informative papers.

I was struck by Martha Riche's observation that the 1990 US Census was seen by Census professionals as the best ever, but by the public as the worst ever. Success is a matter of both technical achievement and perception.

Technical Issues

Pat's paper on maximising coverage tackles the most vital issue of all: the danger is that coverage is seen as something of a dull Cinderella in comparison with other more technically exciting areas.

Some issues for 2001 are:

- The potential benefits of technical innovations in maps, street lists and form design
- The impact of changing the definition of "household" it appears that people will be grouped into larger (and fewer) households
- The need to concentrate resources on the most difficult areas I think that this should be pushed further
- The need to persuade the lifestyle companies not to carry out large scale surveys at the time of the census

Andy's paper on managing the coverage survey sparks the following thoughts:

- Will the survey use different techniques than the enumeration? If not, there is the danger of mere replication. Presumably administrative lists such as driving licenses are non-starters due to the Data Protection Act.
- As it will be voluntary, every effort must be made to maximise the response rate of the survey. I would have a preference for concentrating resources in the most difficult areas, and perhaps reduce the total sample size.
- Matching of enumerated and survey respondent records will not be a clean technical task.

• There should be a quality check too: this may be of modest size, but will be valuable for difficult questions such as income.

In summary, on the technical issues, I hope that we get a really good enumeration, and then season it a little with the CCS. Subsequent evidence from administrative records and rolling forward would be the last resort. Initial results should be checked against 1991 figures to spot any implausible changes. The One Number results should be published - with consistency in totals - and any unadjusted numbers kept firmly out of sight in the research world: there is a great danger of confusing users.

Perception

As well as putting effort into maximising the technical quality, much thought should be put into the issue of perception. The 1991 Census suffered from a constant highlighting of the negative by census specialists in both the universities and the Census Office. The victims were the mainstream users, who were trying to get their organisations to make greater use of Census data to improve decision-making.

In 2001 there needs to be better management of perception. This need not be Orwellian, or in the form of a rapid rebuttal unit, but it should aim to give a better sense of judgement and proportion.

If 2001 turns out to be of similar quality to 1991, but attracts comments that it is a disaster, it will be essential to have a simple message such as:

- Remember that the glass is 97% full rather than 3% empty. 97% really is pretty good.
- Remember that we knew we could never get everybody, We planned for this, and held a large follow-up survey. We've made minor adjustments to the numbers where necessary.
- And, most importantly, remember that the Census is incredibly important for decisionmakers - public services planning schools, hospitals, etc., and business investing in new facilities. Don't knock it.

CHAPTER 20 DISCUSSION: MANAGEMENT ISSUES

Liaison with local authorities to improve census dwellings lists, recruitment of enumerators, output boundaries. The council tax register was suggested as a source of holiday lets, second homes and multiply occupied dwellings. The speakers confirmed that the Council Tax was already used, in particular for multiple occupancy; ONS was open to suggestions for more involvement from LAs, particularly with respect to community groups.

Rachel Leeser suggested ONS could recruit good enumerators via local authorities. They could get the co-operation of Local Authorities, from the Chief Executive down, where local data collection expertise already exists. Bruce Penhale also suggested that LAs mobilise the interest of the local community, recruit enumerators and create the expectation of available jobs during the census. Chris Denham asked for an across the board manner of getting LA involvement as their co-operation is not guaranteed in individual cases. Pat Mann felt that ONS already use LAs and government departments to get the best quality enumerators. However this should be balanced with a concern to provide work opportunities to all.

Chris Denham also felt that LAs will not be needed to be involved as much in 2001 as in 1991. E.g. LAs will not be involved in boundary definition. This statement caused some debate at the session and afterwards. Local authorities felt that LAs should as in the past be involved in suggesting 'soft' boundaries that should be kept to wherever possible, such as housing estates, or other areas that can be defined with local knowledge but not automatically from map features. ONS should tell LAs, via the Local Government Management Board, what electronic format of soft boundaries they can submit, to enable the ONS software to consider their suggestions.

Pay for enumerators. Rob Lewis asked how we pay enumerators peanuts and still expect a good result? Good enumeration is basic. Danny Dorling thought that low rates of pay for enumerators may conflict with the national minimum wage in force in 2001, which would be embarrassing for ONS

Andy Teague felt that although good enumeration is the basis for a successful census, he was not convinced that paying enumerators more will make a difference. With respect to the minimum wage ONS would certainly not break the law. Pat Mann noted that the wages of enumerators are the largest expenditure on the census and the overall budget would be difficult to increase. An extra £100 per enumerator would add £10m to the total cost for example. You would still recruit the same people to do the same basic job. A better approach is to look at simplifying their job and looking at teamwork. **Phone enquiries**. It was suggested that ONS should give phone advice about how to fill out forms, and about the procedure for mailing forms back. For post-back, the forms should be posted back locally for checking and return by the enumerator to improve unusable returns. Andy Teague responded that the precise point to which forms will be returned has yet to be established. Envelopes will be issued with the forms. A helpline number will be available to give advice on form completion.

Tony Champion asked whether enumerators could use the phone to contact people, or in the follow-up? Andy Teague wondered how this might work. The enumerator may not even have a name to look up for a phone number - especially for the people one most wants to contact: the hard to count.

Language other than English on the census form. The speakers confirmed that forms are not issued as a matter of course in different languages, except in Wales. If it is found that noone in a household speaks English, the enumerator has a range of forms available in different languages. ONS will consider keeping records of the issuing of translated documents however enumerators are already being asked to do a lot. What could the specific aim of such records be?

The choice of date for counts. Felicity Andrew felt that the proposal for quoting census counts at mid year rather than as at census night risks undermining public perception of the census. For example, at the end of June students will be at home, whereas on census night they will be at their term time addresses. Andy Teague felt that the choice between census night and midyear is a problem and could undermine the census. All census counts cannot easily be adjusted to 30 June. It is for the customers of population estimates to advise whether they would wish population estimates to be presented as at Census Day or 30 June in 2001.

John Hollis felt that someone within ONS should give a midyear estimate in 2001 in addition to the census, or many people will have to make their own - leading to inconsistent sets of estimates in different organisations

Roma Chappell reported that LAs whom she had asked did want a midyear estimate in 2001 as well as the census results. She in ONS would be interested in hearing other users' views.

How much publicity should be given to plans for census adjustment? It was suggested that talk of adjustment would detract from the obligation to fill in a form. On the other hand, increased co-operation for the CCS would be gained by convincing the public that it was necessary to check up on ourselves, not on them. If we are convinced it needs to be done, then winning public and press and politicians to that view well before the census would avoid political upset after the census.

Quality of 'post-back'. There was considerable concern on this. Jenny Boag reported on a Falkirk pilot of a post-back questionnaire which found a substantial number of people who claimed to have sent back their forms, but for whom no form was received. This represents a possible PR disaster for the census if the same thing happens. ONS responded that the follow-up of post-back questionnaires will be closely managed to ensure this does not happen.

John Boyle reported on the Isle of Man's last census in 1996. Any gain from using post-back was lost because of the follow-up that was required to ensure data quality. The greater the personal contact between the public and enumerators, and the more information an enumerator gives to people, the better the quality of the response.

Andy Teague responded that in an ONS test of post-back it was found that 80% of all returns were posted back in rural areas, and 60% in inner city areas. In most of the country, most of the people will post the forms back, and this will allow more effort to go into difficult inner city areas. We don't want to waste effort walking around compliant areas.

Andy also noted from the pilot that post-back forms had a higher data quality in general. This may be due to some doorstep returners completing forms more hurriedly when the enumerator calls.

Tony Champion suggested a quality check on post-back forms from the CCS pilot.

Information for imputation. Andy Teague confirmed that enumerators will make as much contact as possible, asking and recording the number of people in each household when delivering the forms, to provide an extra check on returns.

Students. Felicity Andrew asked how certain we can be that students will be enumerated at their term time address? How will the ONC target them?

Andy Teague noted that the CCS will be checking between mid May and mid June, and may miss students who return home before then. This stresses the importance of carrying out the CCS as soon as possible after the Census.

ONS will market the census: preferably with as simple a message as possible. ONS will target the message at particular groups such as students, that they need to be enumerated at their term-time address.

On the question of marketing the census, Pat Mann recognised the growth of commercial surveys as a problem. Some surveys even use the word census, ONS does not have the monopoly on that word.

CCS question on migration. Andy Teague confirmed that the CCS is expected to ask a question on address one year ago. In the Hampshire CCS pilot the questions will relate to a Sunday a month previously, as if the census had just taken place.

CCS pilot. There was concern that without the full advertising which will accompany the census, extrapolations could not be made from response to the CCS pilot this autumn. Ian Diamond agreed that the CCS voluntary pilot will have a lower response than the census CCS. In the pilot coverage must be maximised, and extrapolation from the small pilot means that ONS expect 80% response in the full pilot. We must accept that people have a civic duty to comply with the census, but not with the pilot.

Why postcodes? Ian Diamond explained that the CCS will cover postcodes rather than EDs because EDs are too large. There are on average 20 postcodes per ED, but the CCS will on average sample only 4 postcodes per ED to allow a wider coverage.

Measuring dependence between census and CCS. David Rossitter suggested that the CCS could differentiate between the degrees of difficulty of finding different people to help test out the assumption of independence of the CCS and Census. This supported a theme from earlier sessions.

David also wondered if the CCS and the census may conflict for households which are late returning their forms. People completing the CCS who have not yet returned forms may then fail to return the census form in the CCS areas, more than elsewhere. ONS responded that this may happen. A balance needs to be struck on the timing of the CCS - not too early that it unduly affects the census and not too late that circumstances have changed too much or households are unable to recall the situation at the time of census.

Quality check via the CCS. Tony Champion asked whether the CCS will include a quality check on the census, for those variables it uses. Andy Teague responded that quality will be assessed by other means, principally a quality survey after the Dress. Quality will be assessed by other means, principally a quality survey after the Dress Rehearsal.

Tony also suggested that the CCS have an extra more general question about people who would not normally go on the census form, e.g. visitors, and people working away from home.

Further copies of this paper may be obtained from:

CCSR

Faculty of Economics & Social Studies University of Manchester Oxford Road Manchester M13 9PL

Tel: 0161 275 4721; Fax: 0161 275 4722 Email: ccsr@manchester.ac.uk http://les.man.ac.uk/ccsr/index.html

