

Health Data – a health check

A Technical Research Report funded by
Defra

December 2004

AUTHORS

**Institute of Rural Health
Gregynog
Newtown
Powys**

**Centre for Medical Statistics
Keele University
Newcastle Under Lyme
Staffordshire**

Dr Jenny Deaville
Alice Earp
Lesley Jones
Pat Davies
Jane Parry

Professor Gilbert MacKenzie
Nicola Gillon

Department for Environment, Food and Rural Affairs
Nobel House
17 Smith Square
London SW1P 3JR
Telephone 020 7238 6000
Website: www.defra.gov.uk

© Queen's Printer and Controller of HMSO

All enquiries relating to the copyright in this publication should be addressed to HMSO, The Licensing Division, St Clements House, 2-16 Colegate, Norwich NR3 1BQ Fax: 01603 723000 or e-mail: licensing@hmso.gov.uk

Published by the Department for Environment, Food and Rural Affairs. Printed in the UK, December 2004, on material containing 80% post-consumer waste and 20% Elemental Chlorine Free pulp.

Product code PB 10440 ISBN: 0-85521-139-3

CONTENTS

GLOSSARY OF TERMS	4
LIST OF ABBREVIATIONS	5
1. INTRODUCTION	7
2. AIMS	8
3. THE CURRENT RURAL HEALTH AGENDA	9
3.1 Rural deprivation	9
3.2 Equity of access to health care	10
3.3 The impact of rurality on health outcomes	11
4. RURAL HEALTH INFORMATICS	12
4.1 Existing definitions of rurality	12
4.2 The new definition of rurality	13
4.3 The connection with small area statistics	14
4.4 Access to micro data	17
4.5 Measuring health	19
4.6 Aetiological and ecological studies	21
4.7 The nature of spatial risk factors	22
5. RESEARCH METHODS	23
5.1 Overview	23
5.2 Inventory and structures	23
5.3 Sample survey of agencies	24
5.4 Identification of future research at a small area scale	25
6. RESULTS	26
6.1 Structure of the inventory	27
6.2 Overview of the databases	27
6.3 Contents of the main summary table	30
6.4 Sample survey	37
6.5 Identification of future research at a small area scale	39
7. CONCLUSIONS AND RECOMMENDATIONS	39
7.1 Compilation of a comprehensive inventory	39
7.2 Planned research at a small area scale	40
7.3 Assessment of datasets for use in a rural analysis and identification of problems	40
7.4 Recommendations for improving data collection and analysis	41
REFERENCES	
WEBSITE REFERENCES	
APPENDIX A	Background statistical considerations
APPENDIX B	The master table of databases
APPENDIX C	List of contributors to the survey
APPENDIX D	The semi-structured interview schedule
APPENDIX E	Future work at a small area scale

GLOSSARY OF TERMS

Term	Explanation
AZP	Automatic Zoning Procedure - a mathematical procedure for producing geographical zones which are homogeneous with respect to a combination of chosen factors (e.g., population, shape, etc)
Covariate	A variable which may explain some portion of the variation in the response in a statistical model
Criteria for Causality	Criteria (Hill, 1965), which distinguish between casual association and causal association.
Causal & Casual Associations	For example, suppose that, in a statistical model, covariate X explains some portion of the variation in the response Y, but that this association disappears when we take a second covariate Z into account in the model. Then we may conclude that the association between Y and X was casual rather than causal, because it has been explained away by Z. In theory, a causal association should remain even when the influence of Z is taken into account.
Descriptive Study	A study which describes how the response varies in a set of covariates, not necessarily risk factors
Explanatory Study	A study which aims to identify the risk factors for a particular response variable
Explanatory Variable	See Covariate
Health Outcome	The quantity of interest to be measured and possibly to be explained in a statistical model e.g., Incidence, Prevalence, Survival Time, Waiting Time. (Usually measured at the individual level)
Log Odds	The natural logarithm (i.e., \log_e) of the Odds
Odds on an event	If A is an event, then the odds on A occurring is the probability (Pr) of A divided by one minus the probability of A i.e., Odds (A) = $\text{Pr} (A) / [1-\text{Pr} (A)]$
Probability of A, denoted Pr (A)	The chance of A occurring. If A is the event of getting a 'head' in one toss of a fair coin then $\text{Pr} (A) = 1/2$
Putative Risk Factor	A covariate which is hypothesised, <i>a priori</i> , by an investigator, will explain partially or fully the variation in the response
Response or response variable	Health Outcome of direct interest, the variation in which is to be described or explained by a statistical model
Risk factor	An explanatory variable (covariate) which is known to

	have independent explanatory power in a statistical model, i.e., it explains some portion of the variation in the response in its own right and satisfies at least some of Hill's criteria.
Term	Explanation
Statistical Model	An equation in which the response variable on the left hand side is to be explained by a particular mathematical function of one or more explanatory variables (covariates) on the right hand side.
Variation in Response	Variation arises when the individuals studied have different values of the response variable (e.g., Alive/Dead: Waiting times, etc)

LIST OF ABBREVIATIONS

BHPS	British Household Panel Survey
BMRB	British Market Research Bureau
CA	The Countryside Agency
CD	County District
CDSC	Communicable Disease Surveillance Centre at the Health Protection Agency (HPA)
CHI	Commission for Health Improvement
Defra	Department for Environment, Food & Rural Affairs
DfES	Department for Education and Skills
DH	Department of Health
DWP	Department for Works and Pensions
ECT	Electro Convulsive Therapy
ED	Enumeration District
GMS	General Medical Services
HA	(former) Health Authority
HES	Hospital Episode Statistics
HPA	Health Protection Agency
HSE	Health and Safety Executive
ISER	Institute for Social and Economic Research
LA	Local Authority
MHA 1983	Mental Health Act 1983
MHRA	Medicines and Healthcare products Regulatory Agency
MAUP	Modifiable area unit problem
MCMC	Monte Carlo Markov Chain
NatCen	National Centre for Social Research
NDTMS	National Drug Treatment Monitoring System
NFER	National Foundation for Educational Research
NOIDS	Notifications of Infectious Diseases – data held by the Health Protection Agency (HPA)
NOMIS	National Official market Information and Statistics
ONS	Office for National Statistics

PCT	Primary Care Trust
SEAG	Scientific and Ethical Advisory Group
SHA	Strategic Health Authority
UA	Unitary Authority
UKDA	UK Data Archive, University of Essex
VSOB	Vital Statistics Output Branch at the Office for National Statistics (ONS)
YLS	Youth Lifestyle Survey

1. Introduction

The health of rural communities has attracted increasing attention in recent years in the UK, driven by the shift in health policy to focus on health inequalities (Acheson, 1998) and through a significant body of work on the nature of rural deprivation (for example Cloke et al, 1997; Chapman et al, 1998). This closer focus on rural communities, and in particular on the inter-relationship between deprivation and health and well-being has challenged the image of the rural idyll.

There is now a wealth of evidence recognising the nature of rural deprivation, highlighting not only poverty (Shucksmith, 1999), but also opportunity (Chapman et al, 1998) and mobility deprivation (Healthy Norfolk 2000, 1998). It is also recognised that traditional indices of deprivation, such as Townsend and Jarman, are urban biased and that more 'rurally sensitive' methods of data collection and analyses are needed (Deaville, 2001). This focus on rural deprivation corresponds with a growing number of studies at a small area scale, which have identified a range of rural health issues. These include mental health issues and suicide among farming populations (Hawton, 1998), substance misuse among young people in rural communities (Buchan, 2001), physical and socio-cultural barriers to accessing services (Deaville, 2001) and evidence of poorer health outcomes in rural populations for particular diseases such as cancer (Campbell *et al*, 2000) and asthma (Jones and Bentham, 1992). However, with the exception of suicide rates in rural areas, these issues to date have been difficult to identify through the routine examination of national datasets.

This report has been driven by the importance of assessing levels of poor health in rural areas. Defra has a Public Service Agreement (PSA) target to reduce the gap in productivity between the least well performing quartile of rural areas and the English median by 2006, and improve the accessibility of services for rural people. One element of this PSA target addresses access to primary and emergency health care.

The pursuit of these general policy objectives and their longitudinal evaluation requires the availability of data, which are *rurally* sensitive in terms of their geo-spatial definition. A great body of health-related data is collected routinely by Government and other agencies (Office for National Statistics, W1; Cancer Registries, W2) on a regular basis. Whilst official health statistics in the UK have been focussed, traditionally, on mortality data, increasingly the quality and coverage of some types of morbidity data have been improving (National Cancer Plan, W3; TARN, W4).

Many data sets are routinely spatially referenced by geographical boundaries defined solely for administrative purposes (Vital Statistics, W5) while others are referenced by smaller geo-spatial units, which act as building blocks for intermediate structures such as counties (ONS Geography, W6). A variety of other geographies exist and potentially relevant health-related data are stored and routinely reported using these. However, these administrative

topographies may be insensitive to rural policy issues, particularly in relation to the rural health agenda. (ONS Geography, W7)

Thus the importance of framing a satisfactory definition of rurality is clear. A recent report (ODPM, Project Report, 2001, W8) has highlighted the need for revision of the existing definitions of rurality, which are based broadly on urban settlement areas (Census, W10) and rural administrative areas (ONS Geography, W11). When this research project was commissioned work on a new definition was underway (Defra, W12) and it was hoped that this would be available in time to inform the findings of this study. However, at the time of writing, the new definition has not been finalised.

Whatever definition is reached, its utility will need to be evaluated against user needs. In relation to the rural health agenda, major criteria include ease of implementation and usefulness in statistical analysis. It was to examine these and other relevant issues against a backdrop of increasing rural health awareness that this research was commissioned.

2. Aims

The main aim of this research was:

To understand how reliable national datasets are in describing poor health in rural areas.

There were five key objectives:

- To compile a comprehensive inventory of all the datasets collected on poor health and health service use at national, regional and local levels.
- To identify and flag up research planned by health organisations on small area data collection.
- To make an assessment of the suitability of selected datasets for use in a rural analysis, including the degree to which the information can be geographically disaggregated.
- Where there are problems with using the data to describe health levels in rural areas, to identify the nature of these problems
- To make recommendations based on the above as to how Government and others might improve data collection and analysis.

3. The current rural health agenda

The rural health agenda has been increasing in importance in the UK, not only in academia, but also at a policy and service delivery level. In other countries, such as the USA, Canada and Australia, rural health is viewed as a subject in its own right, with well-resourced research and education centres and rural health services. Whilst the scale of rurality in these countries is unarguably greater than within the UK, the proportion of the population living in rural areas in the UK is comparable if not greater. There are currently no specific rural health policies in the UK, but there is increasing convergence of thinking from the health and rural policy sectors. The element of Defra's public service agreement that addresses access to primary and emergency health care in rural areas is a clear example of this.

The increasing attention from policy makers has been matched simultaneously by the research agenda on rural health issues. Research interest has come from a wide range of disciplines including health services research, public health, psychology, rural geography and rural sociology. There is a long history of research from the latter two disciplines into issues such as rural transport, rural economies, rural service provision and rural social capital which provide a wealth of evidence about the wider determinants of health in rural communities. Research into rural issues from the health disciplines has a much shorter history and therefore there is comparatively limited evidence about ill-health in rural communities in the UK.

Recent reviews of rural health research (Deaville, 2001; Higgs, 1999) have identified three key areas of interest to date. These are rural deprivation and social exclusion, equity of access to health care and the impact of rurality on health outcomes. Whilst it is outside the remit of this report to provide a detailed review of current research evidence on rural health, it is relevant to consider the current rural health research agenda under each of these three areas.

3.1 Rural deprivation

There is now widespread recognition of the relationship between deprivation, poor health and reduced life expectancy (Acheson, 1998). However in the past, rural areas have been disadvantaged due to a lack of understanding about the nature of rural deprivation. Historically, indices of deprivation, for example the Townsend Index and the Jarman Index, were implemented across the UK for the allocation of resources and service planning and were widely assumed to be relevant in both urban and rural areas. However, research has highlighted their inherent bias towards urban areas (for example by using variables such as car ownership and proportion of ethnic minorities) and their insensitivity to rural deprivation (RDC, 1998).

There is a wealth of literature supporting the understanding that rural deprivation tends to be 'hidden' because people living in poverty are dispersed across heterogeneous communities. Therefore indices that aim to define 'areas' as deprived are inappropriate. There is also evidence of a culture of

self-reliance and a fear of stigma in rural communities arguably resulting in a low uptake of benefits, thereby hiding the true level of deprivation. Further research is needed into the cultural significance of deprivation in rural areas and the impact that a culture of self-reliance may have on data collected in rural areas.

The Indices of Deprivation 2000 made a step forward in producing national indices that are more sensitive to rural needs. However, there have still been criticisms and they are currently being reviewed. Testing the rural sensitivity of the revised Indices of Deprivation must be high on the research agenda.

3.2 Equity of access to health care

Equity of access to health care is a key policy driver as demonstrated by Defra's PSA targets. Research has considered this topic from two angles. Firstly the equitable provision of services, largely focusing on the allocation of resources; and secondly, the barriers to accessing services.

3.2.1 Equity in the provision of services

There is widespread consensus amongst rural service providers that delivering services in rural areas is more expensive due to increased costs – poor economies of scale, unproductive time spent travelling (by healthcare professionals and patients alike), and additional telecommunication costs (Woollett, 1990). There are also additional costs associated with transport (including higher petrol costs), providing mobile and outreach services, maintaining branch surgeries and accessing training and support. Asthana *et al* (2002) argues that rural communities are disadvantaged in the current system of resource allocation in that it takes insufficient account of the additional costs associated with rural service provision (for example, the current approach is likely to discriminate against areas serving demographically older populations). They argue that NHS resource allocation (at all levels – nationally and locally) needs to be adjusted to account for rurality also known as the “rural premium”.

In terms of future research into analysing the rural cost premium, emphasis should be on research that examines actual costs of providing care rather than on patterns of spending. Woollett (1990) suggests that research into the differential unit costs in urban and rural should be undertaken at regional/local level. This would also have to take into account the needs of the population and the level of spending, as it may be that, due to low provision, rural areas are under-spending and not meeting the needs of their populations.

3.2.2 Barriers to accessing services

Rurality has always added a particular dimension to the issue of access to healthcare services. Evidence of barriers to access is provided in studies of “Distance decay” – where increasing distance from a service has been shown to have a negative impact on uptake – particularly for the elderly, women and low social classes. Deaville (2001) argues that barriers to access can be both physical (distance, lack of transport) and socio-cultural (stigma, self-reliance):

Physical barriers

In rural areas public transport is limited and declining and populations are widely dispersed. There is a high dependence on private car ownership, but often not all family members have equal access to the family car. This can cause problems when trying to access healthcare services, particularly for young people, parents at home with small children and the elderly. Studies into distance decay have suggested ‘cut off’ points beyond which patients will not travel. Research into the optimal and maximum travel times for particular services in rural areas for particular client groups would inform service delivery.

Social and Cultural barriers

Whilst physical barriers to access are the most obvious, socio-cultural barriers are also in evidence. These include the impact of stigma associated with asking for help (particularly with mental health problems) and the lower level of expectation of what services can offer. There is evidence of a culture of self-reliance in rural communities where people may not seek help until illness becomes more serious.

Bearing in mind the distinction between physical and socio-cultural barriers to access, further research is needed into the relative importance of each for different types of service. For example for A&E, social characteristics have limited impact on access in comparison to distance, whilst for mental health services, stigma plays a large role in preventing access.

3.3 The impact of rurality on health outcomes

Research has linked barriers to access as identified above with poorer health outcomes from particular diseases (e.g. asthma, cancer), often because the disease is at a more advanced stage by the time diagnosis occurs. Evidence for the detrimental impact of distance on health outcomes is increasing in the UK and relates to cancer (Campbell et al, 2000), asthma (Jones and Bentham, 1997), uptake of screening (Stark et al, 1997) and thrombolysis (Rawles et al, 1998). Distance decay studies referred to above also clearly demonstrate the decline in the uptake of services with increasing distance. However, little research has tackled the question of whether this is as a result of decreasing need or increasing barriers to access. There is a need to expand the research in this area and look in more detail at whether rurality has an impact on health outcomes, and if this relates to degree of rurality. This question would be

particularly relevant to mortality rates from road traffic accidents, cancer outcomes and 'call to need times' (i.e., the administration of thrombolysis).

The rural health agenda identified here will require a combination of both primary and secondary data. There are detailed methodological issues to be considered when using routine datasets to address these rural health issues. These are considered in the next section.

4. Rural health informatics

4.1 Definitions of Rurality

4.1.1 Urban Settlement Areas (USAs)

Of the two major existing definitions of rurality, that based on urban settlement areas (ONS Geography, W10) has the desirable property of being exogenous to policy issues, since it is based on irreversible urban land use. The recommended definition of an urban settlement area is a piece of land comprised of buildings that are less than 200 metres apart, with a population of 10,000+ persons. There are 1,682 USAs in England. However, in this framework, the definition of rurality is by complementation. Thus land not in urban settlement areas is by definition, *rural*. The virtue of the independence of definition from policy issues has not always been fully appreciated by social, economic and health geographers who have preferred to characterise rurality in other, more systematic, and sometimes policy-related, ways. (ONS Geography, W11).

4.1.2 Rural Administrative Areas

The definition of rural administrative areas (Oxford-CA, 1998, W13) is one such approach in the tradition of classifying administrative structures (e.g., wards) as rural, based on a range of socio-economic characteristics (population density, odds of being economically active, percentage of population in agriculture, forestry and fishing) deemed to define or to be associated with rurality. This approach has the virtue of being nationally consistent, since the classifying variables are usually census related, but it often fails to capture essential geographical features, such as size, form and pattern of the resulting settlement. Moreover, the natural topography is not described accurately.

4.1.3 Evolution of Definitions

While the above urban settlement areas are now regarded as the basis for future work on the new urban-rural definition, it should be noted that a variety of other definitions of rurality exist. It is not the function of this report to review these, but the history of their evolution is interesting, and may be traced in the Review of Urban and Rural Area Definitions (ONS Geography, Project Report, 2002, W7).

4.2 New definition of rurality

These considerations have led Defra (W12) to review the existing definitions of rurality and an updated definition is in the process of being framed. From the outset, a decision was taken to favour an exogenous definition. The proposed new definition builds on one of the two main definitions reported in the Urban/Rural User Guide (W11), which defined urban areas as settlements with a population of 10,000 or more.

The new definition covers a range of geographical levels, the smallest of which is hectare square level. Each hectare square is classified as being urban or rural. Urban hectare squares are those where the majority of addresses fall inside a settlement with a population of 10,000 or more. Rural hectare squares are then further sub-divided into the following morphological types: urban-fringe, small town, village, hamlet, dispersed dwellings. The proposed classification does not currently include a category for those Market Towns with a population of more than 10,000, which, if of interest in a particular application, would need to be recovered from the Urban classification. However, it is likely that such a category will be developed in due course.

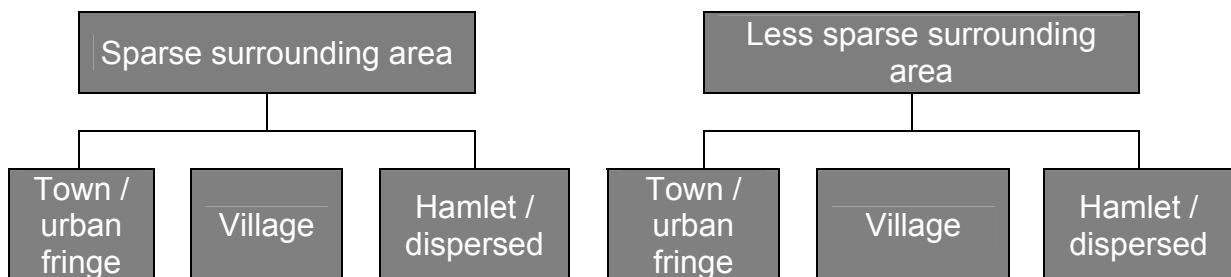
Hectare squares can also be classified on the basis of the following variables:

- (a) **context** (sparsity of the surrounding area) and
- (b) **function** (residential/business mix).

Exact details of the context and function domains have yet to be finalised (Personal Communication, 2002). However, it is anticipated that the context domain will comprise measures of sparsity permitting divisions finer than sparse / less sparse, while the function domain will comprise indices of business mix, based on address classification (e.g., the proportion of businesses associated with agriculture, forestry and fishing).

The proposal is that the classification shown in Figure 1 comprising two broad and six more detailed categories should be offered as the standard breakdown.

Figure 1: Proposed Standard Classification of rural OAs



The hectare square classification is likely to be the preferred option for analysing data as it offers the finest granularity. However, this will not always

be possible due to the risk of disclosure or because data are only available in a geographically aggregated form. Accordingly, larger spatial units such as Census Output Areas (OAs), Super Output Areas (SOAs), electoral wards and local authority areas will also be available.

There are 175,000 OAs in England and Wales containing 125 households on average. OAs are generally smaller than enumeration districts (EDs), which they have replaced. The design of OAs for the 2001 census using generalised, constrained, automatic zoning procedures based on Openshaw's (1997, 1995) proposals was a major task. It has led to the generation of OAs which have more stable boundary definitions over time (e.g., compared with wards), which define a similar number of people, in households with similar characteristics and which are more geographically compact (e.g. compared with postcode districts).

Similarly to hectare squares, each OA is classified as being either wholly/predominantly urban or wholly/predominantly rural. Urban OAs are those where the majority of the population falls inside a settlement with a population of 10,000 or more. Rural OAs are then classified into the following morphological types based on the mix of hectare squares within the OA: urban fringe / small town, village, hamlet / dispersed dwellings. Output Areas have also been categorised on the basis of context and function. Wards and Super Output Areas are classified in a similar way to OAs, except for the fact that the following broader morphological categories have had to be used: (i) urban fringe / small town and (ii) village / hamlet / dispersed dwellings.

4.3 Connection with Small Area Statistics

The new definition highlights the connection with small area statistics (SAS) at least in terms of population numbers, if not in terms of geographical area. This is an area of growing importance in the geographical, epidemiological and statistical communities {ref: ONS, Census Office}. The advent of digitised boundaries, Geographical Information Systems (GIS) and developments in spatial-modelling have led to the realisation of the importance of measuring health data at the small area level (Charles *et al*, 2001).

4.3.1 Rationale

In relation to the rural health agenda, data measured at a small area level provides increased resolution, which may be used to:

- a. describe the demography of local communities more accurately
- b. describe local patterns of disease, including morbidity and mortality
- c. describe local patterns of health need, access and utilisation
- d. identify unusual pattern variation consistent with elevated risk or inequality
- e. provide a basis for generating health research hypotheses
- f. establish baselines for consequential health intervention

- g. provide a means of monitoring and evaluating the effects of health intervention

Thus in principle the concept of measurement at a small area level is sound, as it provides a richer tapestry of information, and data collected at this level is therefore more sensitive to rural health issues. Moreover, modern GIS systems greatly simplify the presentation of this type of detailed information to the consumer. However, measurement on this local scale also generates some important problems.

4.3.2 Confidentiality & Disclosure

Data collected for analytical purposes are subject to legal process safeguarding the confidentiality of the data and anonymity of the individual under the Data Protection Acts (1984, 1998). Public sector guidance on these matters may be found at the Information Commissioner's website (W14). In relation to health, the Information Commissioner exhorts health agencies to observe a *common law* duty of confidentiality, especially in relation to data used for appropriate secondary purposes, e.g. health research or related statistical analysis. When such secondary purposes are warranted, the Commissioner stresses the use of Privacy Enhancing Technology (PET) such as individual record anonymisation and pseudo-anonymisation.

In terms of individual patient records in the NHS, the issues of confidentiality and privacy of the individual have been addressed in the Caldicott Report (1997, W15), which is in broad agreement with the Commissioner's advice, advocating formal agreements between parties exchanging anonymised medical information for valid secondary purposes.

These ethical issues arising in the health setting are not new - they have been handled satisfactorily by census data-managers using a variety of "blurring" devices, for some time. However, the increasing power of GIS systems, computers and statistical packages as well as the increased access afforded by the world-wide-web combines to exacerbate the confidentiality issues. Clearly, as more census data (e.g. age, sex, income, occupation, hamlet) become available with which to cross-classify census individuals, it is more likely that specific individuals can be identified by their neighbours. This is particularly true in rural settings with sparse populations. Thus, for the 2001 census, data-managers employed two "blurring" devices, namely: "population thresholding" and "small cell count adjustment". The population thresholding criteria chosen by the Registrar General for Census Areas Statistics was a minimum of 100 persons *and* 40 resident households. In addition, small cell counts were adjusted (details of the method were not released). These measures are designed principally to preserve confidentiality and prevent disclosure in published tables and the Census Office is not alone in adopting such precautions for data presentation purposes.

4.3.3 Statistical considerations

However sensible the precautions described above they inevitably, impact negatively on the requirements for statistical analysis.

In general terms population thresholding limits the resolution of the smallest areas - perhaps the very areas of greatest interest in a study focussed on rural health issues. Thus, this type of precaution is unhelpful from a logical perspective and, if applied blindly, would limit the usefulness of the census data when addressing the rural health agenda. This does not mean that census data are without value - arguably they are the most vital data resource - providing, *inter alia*, age-sex denominator data for all morbidity and mortality studies, thereby facilitating disease mapping and much other relevant analysis. Accordingly the solution lies in those conducting rural health research and their sponsors entering into formal contractual arrangements with the Census Office in order to receive anonymised, un-blurred data for analysis, but undertaking to observe the thresholding limits when publishing results.

Small cell count adjustment has the effect of impacting disproportionately on (i.e. distorting) the smallest cell counts. Again in the rural health context, this information may be crucial and so it is required without the addition of disclosure control error. The Census Office guidance (W16) warns:

- (a) *"Caution should be taken in interpreting small cell counts as they are impacted to a greater proportion than larger cell counts not only by disclosure control adjustment, but (also) by respondent and processing error.*
- (b) *When calculating proportions, percentages, ratios, from cross-classified or small area tables, the random error introduced can be ignored, except when very small cell (counts) are involved, in which case the impact on percentages and ratios can be significant.*
- (c) *Some small cells (cell-counts) will be altered to zero. Caution should be exercised in deducing that there are no households or people in an area having particular characteristics."*

Thus it is clear that these arrangements are macro in nature. They are not designed to take into account the consequences of the intra-rural topography contingent on the new urban/rural definition. They are also unsatisfactory from an analytical statistical standpoint, which may involve the study of a distribution of rare events in a rural setting, leading to small cell counts (e.g., numbers of accidents or cases of childhood leukaemia). Accordingly, a similar solution of contractual arrangements is required to deal with data subject to small cell adjustment.

Although this discussion has been framed largely in terms of Census practice, the same conclusions should apply whenever thresholding and/or small cell adjustment is encountered.

Other material statistical considerations are concerned with the greater imprecision of estimates of parameters arising in statistical models used to analyse small area geo-spatial data. The increased uncertainty arises due to the smaller numbers in the denominators of the rates being investigated.

A detailed discussion of technical statistical modelling issues is beyond the remit of this report (however, see Appendix A). It should be noted, however, that many modern statistical methods attempt to "borrow strength" from neighbouring areas (Clayton & Kaldor, 1987), thus attempting to make the best use of the data available.

4.4 Access to micro data

These considerations lead automatically for the need in some circumstances for access to so-called microdata - i.e. individual-level data. Many data collection agencies, ONS, Cancer Registries and others appreciate the need for individual level data. The micro data panel at the ONS (Micro Data Panel, W17) is a good example of this type of resource. Requests for access to micro data are received from various sources, including government departments, devolved administrations, independent researchers, and bodies such as local authorities. A senior panel, which reports to the National Statistician, considers all proposals for access and approval is only granted when ONS is certain that the confidentiality guarantee can be maintained absolutely.

The key criteria for assessing whether the confidentiality guarantee can be maintained are:

- data provided by National Statistics must only be used for statistical purposes
- access to microdata, and any subsequent processing, must be lawful and fully compliant with the National Statistics Code of Practice (W18) and the Protocol on Data Access and Confidentiality (W19)
- no statistics will be produced that are likely to identify an individual
- the pledge made and the guarantee of confidentiality given at the time
- of collection will be honoured absolutely
- data identifying individuals will be kept physically secure
- the microdata to be released are to be proportional both to the needs and to the objectives of the research
- access to identified or identifiable statistical sources will only be granted where it is consistent with the aims of National Statistics, as defined in the Code of Practice and the Protocols and with the pledge made to respondents

The existence of this arrangement ensures that critical aetiological research can be addressed in many different arenas. For other reasons, which emerge later, this will be a valuable resource for explanatory rural health studies.

The arrangements in other agencies may be different. Often a Cancer Registry will insist that the research protocol has Local Research Ethical Committee (W20) approval and/or MREC (Multi-centre Research Ethics Committee, W21) approval (if the protocol straddles several LREC areas). In addition, the recipient may be asked to sign a confidentiality agreement with the Registry.

Health agencies in the NHS have been under pressure in recent years to manage patient data and samples more professionally. The implementation of the Research Governance framework in the NHS in England has placed additional formal responsibilities on organisations and investigators alike in relation to the management of personal medical data (NHS Current Code of Practice, W22). The rules for managing personal medical data and maintaining confidentiality within the NHS are enshrined in the Caldicott Report (Caldicott, 1997). However, while the Caldicott report set up Data Guardians and listed justifiable reasons for the transfer of confidential medical records on a 'need to know' basis, the standpoint was entirely introspective, and Caldicott did not address the issue of inter agency exchanges for research purposes.

The full implementation of all of these arrangements in NHS agencies is far from complete and it will take some time for staff to become experienced in their operation. One view among the statistical community is that the net effect of these measures has been to create unnecessary and artificial barriers to data exchange (e.g., some NHS Trust R&D handbooks specify that personal medical data must *never* be exported). Inter-agency personal data transfers fall within the scope of the Data Protection Act and the intention of this Act is to regulate, but not thwart, transfers for legitimate research purposes. The need for a standard interchange protocol and a centralised administration governing the transfer of anonymised individual-level data on a national basis is clearly overdue.

The position in Ireland, for example, is quite different and perhaps lessons could be learned in the UK. There, the Statistics Act 1993 (W23) gives wide powers of access to the Director General of the Central Statistics Office and his staff. In relation to accessing data held by Public Bodies the Act is specific:

(1) For the purpose of assisting the Office in the exercise of its functions under this Act, the Director General may by delivery of a notice request any public authority to —

- (a) allow officers of statistics at all reasonable times to have access to, inspect and take copies of or extracts from any records in its charge, and
- (b) provide the Office, if any such officer so requires, with copies or extracts from any such record, and the public authority shall, subject

to subsection (2) of this section, comply with any such request free of charge.

(2) Subsection (1) of this section —

- (c) shall not apply to records pertaining to a Court, the Garda Síochána, the prison administration or the Ombudsman or any of his officers;
- (d) shall apply to medical records which are not publicly available only with the agreement of the Minister for Health;
- (e) shall in all other cases have effect notwithstanding anything contained in any enactment other than provisions for the protection of public order or the security of the State.

Subsection 2 (b) shows the need for a National Expert Panel advising the Minister for Health and the Director General.

Any similar legislation in the UK should apply to all anonymised court, police and prison records. There is, for example, ample evidence of poorer health among prison populations.

4.5 Measuring Health

Whatever geo-spatial resolution is adopted is likely to be subject matter dependent and we turn now to consider the broader issue of measuring health and the merits of particular types of study design. The World Health Organisation defines health as: "*A state of physical, mental and social well-being and not only the absence of disease and disability*". This is a very broad definition requiring the measurement of many medical, social and economic indicators to determine whether it is satisfied.

4.5.1 Orthodox Measures

Traditionally, the health of any population may be gauged broadly by considering measures of ill health:

- (a) **Mortality Rates** - numbers of deaths from all or specific causes per unit of population
- (b) **Morbidity Rates** - numbers of cases of (non-fatal and possibly recurrent) illnesses per unit of population.

The measurement of mortality has been complete and systematic since it was introduced by William Farr^{*} in 1841 and has been conducted annually since, via returns to the Office of the Registrar General. The virtues and limitations of the system are well understood by the medical and epidemiological communities engaged in health measurement.

^{*} Compiler of abstracts at the Registrar-General's office, 1838–79, FRS 1855.

However, the measurement of morbidity (e.g. breast cancer, spells off work, bouts of influenza, cardiac arrests, exposures to HIV infections, back pain) is relatively incomplete on a national scale and such data as exist are much more ad-hoc, with respect to coverage and representativeness. Traditionally, the measurement of ill-health has been geared more to making a single set of measurements on each individual (i.e., cross-sectional data such as age, sex, presence or absence of a condition) and is most frequently related to episodes of care in an NHS setting. However, many types of morbidity are recurrent and their measurement necessitates longitudinal (i.e. repeated) observations on the same individual. Primary care is the obvious setting for the collation of this data. However for several reasons, including incompatibility of computer software, confidentiality issues as outlined above and expense, these data are not collated at a national level. Reliance is placed instead on sample surveys (e.g., ELSA, W24) and selective indirect methods (e.g., Primary Care Research Networks, RCGP, W25). Perhaps the exception to this rule is the Cancer Registration Service where ascertainment of cases may be largely complete on a national basis, but socio-demographic characteristics and details of treatments received are frequently incomplete (Gillon & MacKenzie, 2003).

4.5.2 More indirect measures

Government reform of the National Health Service, improved Clinical (W26) and Research Governance (W27), coupled with the rise of Information Technology have facilitated the systematic collection of patient information at the point of service (e.g. GP Consultation, Hospital Admission and Community Services). It may be thought that the accumulation and integration of such information will prove to be a rich source of research data. However, in general such data are not collected for specific research purposes but rather for administrative purposes and this limits their usefulness in the health research arena. A rather obvious example is the way in which hospital data are accumulated - on an admission or by completed consultant episode basis, when existing hospital "information" systems make it difficult to assess the cumulative history of any single patient through the system over time (Li & MacKenzie, 2001). Curiously, individual patient outcome over time appears not to be a priority NHS indicator of hospital performance at a time when many hospitals are under-pressure to discharge early.

A further complication relates to completeness of ascertainment, since many types of minor morbidity (e.g., trauma) may never come to the attention of relevant service providers. Other indirect measures of health relate to inequality of access to existing facilities, which may be predicated on a set of underlying socio-economic and demographic factors, as highlighted in section 3.3.2, necessitating their measurement in the population over time. (Black Report, 1980, W28; Acheson Inquiry 1997, W29; DOH Tackling Health Inequalities, W30).

4.5.3 Measuring Rural Health

The measurement of rural health falls squarely within the framework described above. However, whilst the measurement of mortality is unlikely to be influenced directly by geographical factors the measurement of morbidity is a different matter.

It is not difficult to see how the interplay of issues highlighted in the rural health agenda above might influence the measurement of morbidity. Concentrations of key medical and clinical resources in urban areas, long travel times and socio-cultural factors all have an impact. A key example of the impact of these factors is the accurate measurement of accidental injury, particularly related to agricultural activity. Greater distance to the local casualty department mean that minor injuries are more likely to be seen in primary care in rural areas (Evans, 1998) or even not seen by an NHS provider at all, especially if there is fear of repercussion from being reported to the Health and Safety Executive. At a national level, studies of accidental injury more frequently rely on data from accident and emergency units rather than from primary care, due to the relative accessibility of data. Consequently any geographic analysis of injury seen in accident and emergency units is likely to be under-representative of minor injuries taking place in rural communities.

4.6 The respective roles of aetiological & ecological data in addressing rural health issues

In a given research problem with a rural dimension, a particular causal hypothesis will have been formulated and a study designed to investigate it. Appendix A provides a detailed background into the essential logical characteristics of the classical types of study, which are available to investigate such hypotheses and, the type of corroborative evidence they provide when conducted properly.

We learn there that, *inter alia*, the classical *aetiological* epidemiological study requires data measured at the individual level rather than (merely) on the spatial units in which the individuals live (ecological data). Thus, aetiological epidemiological data are quantitatively richer and qualitatively distinct from ecological data.

The requirement of using data at an individual level arises because Hill's Criteria of Causality (Hill, 1965) are based on a framework of logic, built up from the Henle-Koch postulates (Evans, 1978) for disease causation, as applied to *individuals*. There is no corresponding framework for spatial units and, indeed, even if one existed it would not be relevant to individuals.

The distinction is an important one because it limits the role of ecological data to a description of the spatial units and highlights the logical difficulties of attempting to explain, in aetiological, i.e. causal, terms, spatial patterns of disease using ecological data.

This logical difficulty is termed the "Ecological Fallacy" (Robinson, 1970) but we shall refer to it as the "Aetiological Fallacy", because of the aetiological constraints implicit in ecological studies of health. Thus, the "*Aetiological Fallacy*" is arguably more serious than the modifiable areal unit problem (MAUP), in which associations in space may be generated simply by the particular choice of spatial unit for analysis (Holt, 1996). In this case, the choice of another type of spatial unit may lead to a different spatial pattern of disease and thus to a substantially different interpretation.

Thus we may summarise by remarking that, in general, ecological data can answer "where" questions (MAUP apart), but cannot address "why" questions. Accordingly we may associate the aggregated data, which typify ecological studies with *description*, while realising that individual level data are required for *explanation*.

These logical constraints attend the analysis of all ecological data and consequently impact on rural health studies where individual level data will be required to address aetiological research questions.

4.7 The nature of spatial risk factors

In relation to health studies the nature of putative spatial risk factors must be taken into account. There are apparently few naturally occurring spatial risk factors in comparison to those created by man's interaction with the environment. We may consider radon in rock (Green *et al*, 2002) and arsenic in well water (Howard, 2003) as two obvious examples of *naturally* occurring spatial factors that adversely affect health. However, we should consider water pollution from disused mine workings and particulate radioactive contamination as artefact.

In many cases we may anticipate that a particular spatial pattern may arise simply because we have failed to account for some other, more fundamental, risk factor which follows a similar spatial pattern for transparent reasons. For example, poor breast cancer survival rates may follow the spatial distribution of the elderly, which may be coastal or show apparent clusters where there is a concentration of nursing home accommodation.

Thus, allowing for the influence of age in an appropriate statistical model will have the effect of abolishing or explaining the (spurious) spatial association. This is why the 9th criterion of causality is added to the eight proposed by Bradford - Hill (Appendix A). It is only when a spatial pattern cannot be explained by potentially confounding factors that it merits further investigation. And so might many *apparent* urban/rural effects be explained in the same way.

5. Research methods

5.1 Overview

Our plan was to identify existing data sets considered to have potential relevance to the rural health agenda and to focus attention on the core databases within this set using standard selection criteria. The core databases were structured according to key health themes and an annotated master table was created. The contents of this table were classified according to several criteria pertaining to their usefulness in the light of the new definition of rurality.

A random sample was drawn from the original inventory and the various agencies involved were identified. The sample of agencies was augmented by additional purposive selection of some additional agencies, which the random selection had missed. Discussions with the augmented list of agencies were held and these took the form of a qualitative research survey involving semi-structured interviews designed to elicit information on topics germane to the objectives of this study.

The research survey was conducted over a 6 month period (1st September 2003 to February 2004 inclusive) by the Institute of Rural Health and the Centre for Medical Statistics (Keele University). The departure from the original schedule was occasioned by the delay in obtaining the new definition.

5.2 Inventory & structures

Compilation of the inventory was the essential first step and was commenced at the outset of the study.

Potentially relevant datasets were identified using a web-based search conducted by the librarian at the IRH (JP). The number of datasets for consideration is potentially very large in this project. Accordingly, the following set of inclusion criteria were employed to narrow the search:

1. Datasets in publication since 1997 (includes datasets begun prior to 1997)
2. Datasets whose contents are made available either in electronic (via the web) or printed format (e.g. book published every year)
3. Datasets for England – this may include datasets covering the UK, or England and one other country.
4. Datasets dealing with health, illness, health service use and/or well-being/quality of life

In total 153 datasets were identified. The entire inventory was scrutinised by a research assistant (NG) to identify the core datasets and organisations involved. Where possible the core dataset was traced and key items of information concerning the dataset were obtained from the organisation responsible for its acquisition and/or maintenance. The following properties were of particular interest:

- (a) Type of data - Census or Survey
- (b) Spatial resolution
- (c) Ecological or Aetiological data
- (d) Ease of Implementation of the New definition

The inventory was re-structured in order to eliminate duplicates, references to the same datasets, or families of datasets, and an annotated master table was created with respect to the properties (a) - (d). The full master table is included in Appendix B.

5.3 Sample Survey of Agencies

5.3.1 Selection of sub-sample of agencies

To obtain a greater insight into the rural relevance of datasets in the inventory a sub-sample of 31 datasets were selected via random (25) and purposive sampling (6) and telephone discussions held with a key contact for each dataset (the list of contributors is given in Appendix C). Table 1 shows the final list of 26 agencies contacted successfully.

Introductory letters were sent to the selected agencies requesting a telephone discussion and then followed up with contact by a Researcher (LJ and PD) in order to arrange a suitable time for the discussion.

There were no refusals to participate. In two cases, one individual was responsible for 2 datasets and so the discussion was held once with each of these individuals covering the datasets of relevance. In one case the database had been concluded and in 2 cases it was not possible to make an appointment to conduct the discussion.

5.3.2 Structure of the discussion

In order to structure the discussions a set of discussion points were drawn up (see Appendix D), agreed with Defra and used by the Researcher in each contact. Written notes were taken at the time of the telephone discussion, which were then typed up soon afterwards. The discussions were undertaken by two Researchers at IRH (LJ and PD).

5.3.3 Analysis of data

The transcripts of the discussions were collated and an initial sorting of the data and thematic analysis was undertaken by one of the Researchers who conducted the telephone discussions (LJ). Another Researcher (AE) who had not been involved in the telephone discussions then undertook a more in-depth analysis.

5.4 Identification of future research at a small area scale

Information on future potential sources of data at a small area scale was obtained through the following methods:

- i) Trawl of the NHS National Research Register
- ii) Networking with contacts. An abstract of the research project with a request for information on small area scale research was circulated to the following:

- IRH membership (100 contacts)
- Rural Health Forum Exchange (125 contacts)
- Rural Health Forum Advisory Group (19 contacts)
- Defra rural interest group
- Strategic Health Authorities (22 contacts)
- NHS Confederation (48 contacts)
- Rural Community Councils (33 contacts)
- Rural Primary Care Trusts (65 contacts)

Information was collated and entered on to a database for analysis.

Table 1 Sub sample of datasets for the sample survey

	Dataset Name	Organisation(s)
1.	National database for Primary Care Groups and Trusts	National Primary Care Research and Development Centre, University of Manchester
2.	Breast Screening Programme Statistical Bulletin	Department of Health
3.	Children Accommodated in Secure Units in England & Wales	Department of Health and Department for Education and Skills
4.	Children Adopted from Care in England	Department of Health and Department for Education and Skills
5.	Community Care Statistics: Adult Home Care	Department of Health
6.	National Drug Treatment Monitoring System	Department of Health
7.	Guardianship under the Mental Health Act 1983	Department of Health
8.	Hospital, Public Health Medicine and Community Health Services Medical and Dental Staff in England – from Annual NHS Staff Censuses	Department of Health
9.	Personal Social Services Expenditure	Department of Health
10.	Prescription Cost Analysis for England	Department of Health
11.	Health Survey for England	Department of Health, NatCen, Department of Epidemiology and Public Health at University College London
12.	National Patients Survey	Commission of Health Improvement
13.	Hospital Waiting Times	Department of Health
14.	Fatal Injuries	Health and Safety Executive
15.	British Crime Survey	Home Office, BMRB International
16.	British Household Panel Survey	Institute of Social and Economic Research, Institute of Fiscal Studies
17.	Mental Health of Carers	Office for National Statistics, Department of Health
18.	Probation Statistics	Home Office
19.	Birth Statistics	Office for National Statistics
20.	Better or Worse: a longitudinal study of the mental health of adults living in private households in Great Britain	Office for National Statistics
21.	Contraception and Sexual Health – module from the ONS Omnibus Survey	Office for National Statistics
22.	Cancer Registrations	Office for National Statistics
23.	Housing Statistics	Office of the Deputy Prime Minister
24.	Attendance Allowance	Department for Work and Pensions
25.	Communicable Diseases	Health Protection Agency
26.	Agricultural Census	Defra

6. Results

We present the results in three sections: first a discussion of the inventory of datasets; second a summary of the findings from the interviews with a sub-sample of agencies; and finally the results of the search for future and planned research at a small area level.

6.1 Structure of the inventory

The original inventory, which employed broad search criteria, comprised 153 dataset entries. Many of these were simply web-page references to statistical bulletins or annual reports summarising the contents of national datasets for a particular time period. Using this inventory, our first task was to identify the underlying core datasets and the organisations responsible for their acquisition.

The 153 original inventory entries were reduced to 120 in the following way. Eleven of the inventory items were references to datasets already included elsewhere in the inventory. For example some pointed to the same series of a report, but perhaps for a different year. Another 11 were not available publicly for 3rd party research, except in the prescribed format of a published report. Such "data" may form useful background material but in general are too limited for rural health research. Three datasets were only available at a national level and one data-set comprised data aggregated to Government Office Region level, which is not suitable for a rural health analysis. The coverage or sample used for three of the datasets was thought to be too incomplete or too small to represent rural areas implicit in our brief. One dataset only covered Scotland, another was no longer available, while a third referenced datasets already in the inventory. One of the inventory items was for a national development programme but detailed no specific datasets, other than those that were already included elsewhere on the inventory. Finally for another dataset no contact details could be found.

On closer examination we found that the remaining 120 inventory items could be reduced further - by grouping some datasets - to yield a core of 103 dataset items. We have listed and annotated these in the master database table (Appendix B).

6.2 Overview of the databases

A summary of the major themes is shown in Table 2. The 103 datasets were first classified under two main headings: "Directly health-related" (16 subgroups) and "Wider determinants of health" (4 subgroups).

TABLE 2 Classification of the Themes in the 103 core Data-sets	
DIRECTLY HEALTH-RELATED	WIDER DETERMINANTS OF HEALTH
Vital Statistics	Child Care
Population and Internal Migration	Crime
Comprehensive General Practice/PCT Databases	Labour Market/Unemployment
NHS Performance/Waiting times	Other Socio-economic Data
Health Service Staff Characteristics/Recruitment	
Cancer	
Mental Health	
Other Long-term Illness/Disability	
Work-Related Disease	
Accidents	
Infant Feeding	
Sexual Health	
Smoking, Drinking & Drug use	
Other Disease/Illness Related Statistics	
Other Health Services Statistics	
Other National Surveys with Health-related Data	

Directly health-related datasets include data, which are associated with particular health conditions or illnesses, or with the health services. However, we also included vital statistics and population information within this group, as this information is needed to monitor the general health of the population, and to calculate rates of health factors in the population. The “wider determinants of health” datasets contain information, which may indirectly impact on health (or have a ‘knock-on’ effect). In the main, the information covered in these datasets concern employment, childcare, income, accommodation and the impact of crime.

It is not straightforward to classify datasets exclusively under one of the two major headings or under a single subgroup. We have tried to structure the table so that it reads coherently, by only entering each dataset once under the ‘most relevant theme’, but clearly some of the datasets could have been placed in several subgroups. This is particularly true for the national census data and the majority of the national survey data-sets that we have covered, which comprise a wide range of different variables, some of which come under the “wider determinants of health” while others are directly health-related. Note that the marriage statistics dataset has been included under ‘vital statistics’, which is classed as a “directly health-related” theme. Vital statistics are often used as key indicators of the general health of a population, and it seemed most sensible to include marriage statistics with the rest of the vital statistics.

To refine the picture we classified the 103 datasets into those, which contained only variables that were directly health related (40%), those which were only related to the wider determinants of health (25%) and those which contained both types of variables (36%).

The 'directly health-related only' group were mainly compiled by the Department of Health to monitor the activity and performance of the wide range of health services available in the NHS. The types of datasets held in this group ranged from (a) aggregate data collected for a few specific factors, such as the numbers of first and initial contacts (by age group) with the chiropody services in one financial year within the NHS, to (b) very comprehensive databases containing individual patient level data on numerous health variables, such as the General Practice Research Database (GPRD), The Health Improvement Network (THIN), Hospital Episode Statistics (HES) and the Trauma Audit and Research Network (TARN). A large proportion of the dataset items (36%) hold variables of both types. The datasets classified as 'wider health determinants' mainly consist of information on employment, income, household composition, accommodation, child-care and impact of crime.

Most of the national surveys within the inventory are included in this 'both' group, and collect broad information on most of the factors listed above. However, with the exception of the Health Survey for England which collects comprehensive information on health, the health data covered in most of the national surveys are quite limited – being related to general health, recent use of health services, opinions on the health services and the effects of illness on work and other activities.

However, the fact that wider socio-economic variables can be linked directly at an individual level to these general health variables makes these data a valuable resource. Of course, the National Census data are also included in this group. There were more health variables covered in the latest 2001 census than in previous censuses, including self-reported general health in the previous 12 months, long-term limiting illness and the provision of unpaid care. The census data are held on the Neighbourhood Statistics section of the Office for National Statistics website, where other data are also deposited including benefit claimants and employment estimates, education data, selected hospital episode statistics, housing and deprivation indices.

Only 25 (24%) of the datasets included did not contain some directly health-related information. Most of these datasets dealt with non health-related benefits (including unemployment benefits), crime (including the British Crime Survey) or the care of children by local authorities.

6.3 Contents of master table

In the master table (Appendix B) we provide an annotated list of the 103 core datasets. The annotation includes a brief description of each in terms of the type of data collected, its available spatial resolution, and the potential that each dataset appears to have in relation to implementing the new definition of rurality. Some additional explanatory comments appear in the last column.

6.3.1 Types of Data - Census or Survey

Each dataset was either a sample survey or a census (100% sample). We have looked at a number of the main national surveys, including the Labour Force Survey, the Health Survey for England and the ONS Omnibus survey in detail. For surveys, we have included the approximate sample size used for the latest survey for which data are available.

The data for most of the national surveys are deposited for public access on the UK Data Archive at the University of Essex (www.data-archive.ac.uk), usually a year or two after the fieldwork was carried out and the initial analysis has been completed. Other datasets, which use a sample methodology include some of the benefit statistics from the Department for Works and Pensions (DWP) (e.g. Bereavement Benefit, Maternity Allowance) which use a 5% scan of all claimants from the corresponding computer systems. The majority of the benefit statistics are now produced from a 100% scan, however, or are moving towards using this in the future.

For any datasets using sample methodology, the data analysis needs to take into account the method of sampling that was used, and the resulting coverage in space and time of the dataset. For example, certain social groups may have been over-sampled to obtain sufficient numbers for analysis (i.e. the proportion of representation in the sample for these groups is larger than the overall population proportion), which would need to be taken into account in the weightings used in the analysis. On the other hand if the sample is nationally representative then this will generally mean that the 'rural' group will be much smaller than the 'urban' group of individuals within the sample. In this case it would need to be checked whether the 'rural' group was large enough for the analysis in mind. The type of sampling used for the different datasets is not presented in detail in the interests of brevity. However, sampling methods are discussed along with other technical details in metadata provided by the organisations, which are publicly available on the web for all of the datasets listed. A final caveat is that the data may be perfect in all respects but simply out of date.

There are a number of comprehensive databases included in the master table, e.g. the General Practice Research Database (GPRD), The Trauma Audit & Research Network (TARN), which hold patient level data collected from general practices or hospitals that agree to participate. So in these datasets the coverage of data depend on the level of participation of the relevant health organisations. The numbers of participating organisations for these databases is increasing all of the time.

The remaining datasets are generally made up of censuses, which in theory should give 100% figures. Much of the census data are collected regularly on statistical return forms. Whenever possible we have noted the reference numbers of the relevant forms used in this case. The Department of Health (DH) use statistical returns to monitor the activity of the different health services provided by the NHS, whilst the Home Office monitor the activity of the police forces and probation services in a similar way.

6.3.2 Spatial Resolution

The 'Available spatial resolution' column in the master table details the smallest geographical unit for which data can be released. This is one of the key factors determining whether a rural/urban classification could be usefully attached to the data. Confidentiality is obviously an important determining factor on the spatial resolution of publicly released material, and one which all of the organisations take very seriously. Typically, the spatial resolution of the dataset is fixed so that the data are not disclosive, and so that individuals cannot be recognised from any breakdown released. The majority of the datasets are only available for release at quite a high geographical level, such as PCT/NHS Trust or local authority/unitary authority level (LA/UA). Some of the datasets consist of aggregate data at this level which cannot be broken down any further to lower level for example the Hospital Activity Statistics (under the 'NHS Performance/Waiting Times' theme) are produced from a number of statistical return forms which can clearly only be broken down to NHS Trust level.

Other national datasets are obtained by collecting anonymised, individual-level data covering a variety of relevant variables, from the different local organisations such as PCTs or local authorities. For example, for the 'Guardianship under Mental Health Act 1983' dataset (under the 'Mental Health' theme in the master table) anonymised data are collected annually by the Department of Health from each local authority. Therefore, although the dataset contains individually measured data (rather than aggregate data) for a number of key variables, publication is at the local authority level. For the original purpose of such a dataset, local authority level data are sufficient, but for use in rural health research, this level is too broad brush. In this situation, it may be possible to obtain geographical information at lower levels from the local authorities directly, in order to implement a rural classification at lower levels. Of course, to use such data, the proposed research would need to be formally approved by the appropriate official bodies, and the maintenance of confidentiality guaranteed by formal agreement. .

Many of the national survey data can only be released down to Government Office Region level – due to confidentiality and the sample sizes used. An exception to this is the British Social Attitudes Survey, for which data are available down to ward level. However for the British Social Attitudes Survey, the questions on health are only asked to a random two-thirds of the total sample and this may be too small for useful urban/rural analyses.

Some datasets are available at lower resolutions, notably the Census information which is available at census output areas. However, the blurring methods used by the Census Office for areas with small counts must be kept in mind when analysing data at this level. Of course, rural areas will be particularly affected by small counts. Most of the birth and mortality statistics are available at ward level, except for particularly sensitive data that may be more identifiable such as for stillbirths or infant deaths, which are only available to local authority levels. A number of different benefit statistics data are available at ward level too, when the numbers of claimants involved are large enough to prevent disclosure. However ward level benefit claim number is rounded to the nearest 5.

If a particular research protocol is approved by the relevant authorities and confidentiality agreements are set up, it is possible to obtain lower level data than are generally released into the public domain for some data-sets. For example it may be possible for certain researchers to obtain individual data from the cancer registrations databases, including postcode information, for statistical analysis. In this case, it would clearly be easily possible to add rural classifications at an individual level. The level at which the results of any analysis of such data could be published would have to be agreed so that confidentiality was not compromised.

A further example is that although the national survey data can generally not be released at low geographical levels, it is possible that, given a rural classification, the survey research team could attach this using postcode information, within an 'in-house' safe-setting, for a reasonable charge, and then release the data with the rural classification attached for approved research. In this case, the (location specified) geographical resolution of the data would remain quite large, but the rural classification could be used to compare different rural/urban area types without specifying particular locations in great detail. The release of such data would be subject to whether any breakdowns using the attached rural/urban classification would be disclosive.

In other cases, individual organisations will conduct a full analysis of the data 'in-house' to ensure confidentiality. For example, if a research protocol is approved, TARN team will carry out research or allow other researchers to work within the TARN team, but data would not normally be released outside the aegis of the TARN team.

6.3.3 Potential for new Definition of Rurality

We have tried to form an indicator of the ease or difficulty of including the new rurality classification within each of the 103 dataset items in the master table to take account of the points discussed above. In the column named 'Potential for using new definition of rurality' we have used symbols to indicate this, which are explained in the key below:

Figure 2 Key to symbols in the master table

- ✓ Data can be disaggregated and published at areas smaller than LA/UA or PCT level.
- ☑ Data can be disaggregated to smaller areas than LA/UA and PCT for attachment of rural/urban indicators or for statistical modelling etc., but data cannot be published at these smaller areas.
- ☐ Data can only be disaggregated at the most to LA/UA or PCT level and rural/urban indicators cannot be attached at lower levels.
- ✗ Data are unlikely to be analysed meaningfully using rural/urban distinctions.

Data-sets with the greatest potential in using the new rurality definition are those that are allocated either of the first two symbols in the key.

The first symbol ✓ indicates that the dataset can be publicly published at quite low levels, such as ward or even output area level, and for which rurality classifications could be applied readily. For this type of data, it would be fairly easy to add a rural classification, whenever one was needed, even if such a classification was not added as standard within the dataset.

The second symbol ☑ indicates datasets for which data cannot be publicly published at low levels, but for which it is possible to attach rural indicators to lower level data such as postcodes, and then use the rural classification within the statistical analysis, even if the area locations cannot be identified at a low spatial resolution. An example of this is when a survey research team could add the rural classification to the dataset and then release the data with the classification on an ad hoc basis for an approved research project. Of course, if such postcode information were available, then for these datasets it would be possible to add a rural classification as standard within the dataset, if it was found to be useful. This symbol was also allocated to datasets for which it may be possible to obtain full postcode information for approved research, e.g. hospital episode statistics data, for which full individual-level data may be obtained, but which are required to be released at a higher geographical level.

The third symbol ☐ indicates datasets that currently cannot be disaggregated to lower levels than PCT or Local Authority. The resolution of this geography is not of sufficient sensitivity to be used in a rural/urban analysis. This symbol was allocated to aggregate datasets and also datasets that consist of anonymised individually measured data, at this level. In the latter case, although the centrally held national dataset cannot be broken down to lower geographical levels, it may be possible to recover postcode information

directly from the organisations which provide the data, such as the local authorities, if the research was approved.

The fourth symbol × indicates datasets for which a useful rural/urban analysis is unlikely to be possible. Only 2 datasets in the master table were allocated this symbol. These datasets might well have been rejected initially. However, they are good examples of datasets that seem to be unsuitable for rural health research as they stand. The first entry allocated this symbol is the General Ophthalmic Services datasets on the Ophthalmic practitioners. More than half of ophthalmic practitioners hold contracts with more than one PCT. It is therefore not easy to break the ophthalmic workforce down geographically, and if it were possible, it could only be done at quite a high level. The second table entry is for the data collected by the Department of Health and the Department for Education and Skills on Children Accommodated in Secure Units. Here the data can only be broken down to secure unit level, for which there are only 30 in England (1 in Wales), and no data on the child’s former residence are collected. Therefore, again, these data can only be broken down to a high geographical level, which is unlikely to be useful in rural health research.

Table 3 below summarises the distribution of these symbols that have been allocated to the different dataset items in the master table, again according to whether the information held is directly health related and/or concerns wider determinants of health.

From Table 3 we can see that the datasets which have the highest proportion of ✓ and ☑, indicating a higher potential in being able to include the new rurality definition for useful rural health research in the group of datasets which include both directly health-related variables and variables measuring wider determinants of health. This group of datasets consists of most of the national surveys, the national census and other Neighbourhood Statistics data, and some of the health-related benefits such as disability allowance, attendance allowance and incapacity benefit. Of course, as discussed earlier, the coverage of any samples used needs to be evaluated to check whether it is sufficient for a rural health analysis. Also the sampling methods used, if applicable, need to be taken into account in any data analysis.

Table 3: Distribution of the symbols indicating the apparent potential of data-sets in using new rurality definition by the type of data held

		Directly health related		TOTAL
		Yes	No	
Wider determinants of health	Yes	37 (36%) = 8 ✓, 20 ☑, 9 □, 0 ×	25 (24%) = 7 ✓, 7 ☑, 10 □, 1 ×	62 (60%)
	No	41 (40%) = 3 ✓, 7 ☑, 0 □, 1 ×	–	41 (40%)
TOTAL		78 (76%)	25 (24%)	103 (100%)

So out of the total 103 dataset items detailed in Table 3, there are 18 ✓, 34 ☑, 49 ☐ and 2 ✕ symbols.

The vast majority of the ☐ symbols, indicating a lower potential within the corresponding datasets for including the new rurality definitions, are in the group of datasets, which only contain directly health-related variables. These datasets are, in the main, statistics compiled by the Department of Health, which monitor the activity and performance of health services, mainly at PCT or NHS Trust level. These data are mainly collected as aggregates, and so cannot be broken down to a level that is rurally sensitive.

6.4 Sample Survey

The statistical description of the data in the master table (previous section) was augmented by direct contact with agencies involved using a semi-structured interview approach. The discussions with agencies were held around three key areas: first, the context of the datasets (for example the purpose and audience of the datasets, sampling methods and age-sex structure); second, data quality and confidentiality issues; and finally the rural relevance of the data. The following discussion is structured around these three main areas.

6.4.1 Context of the datasets

The objective for compiling the various datasets varies. However in general information is gathered for one or more of the following four priorities: to inform policy; to assist in decision making about a particular service; to monitor progress and trends; or to identify issues of concern.

Government officers are key users of the information that relates directly to the objectives of many of the datasets. However in most cases the information is published and available to the general public via the Internet or through direct contact. 65% of the agencies spoken to cited the public as one of the main user groups. Academics and researchers are also a key user of data.

More than half of the databases were relevant to England only. However, a large number were relevant to England and Wales, thus indicating that devolution in Wales has not yet altered the requirements of national datasets. The 'edge effect' (as a result of geographical boundaries in service provision) can affect the data in that where data are collected only for England, Scottish and Welsh residents may be included due to cross border movement for accessing services.

Sampling methods vary according to the objective of the dataset. Datasets range from 100% sampling of all individuals with a specific disease (for example the Cancer Registry) or waiting for contact with secondary care (Hospital waiting lists) to random sampling for surveys (for example CHI Patient Survey). The Agricultural Census (Defra) is an example of a dataset

where some data are imputed (calculated from previous years averages rather than inputting real data).

6.4.2 Data quality and confidentiality

Data quality is taken seriously by all agencies involved in the discussions. Some 6 out of 26 agencies felt that the standard of their data was very good or extremely high. Most agencies are convinced that the data quality is good and 90 to 95% was felt to be a good standard. The majority of agencies have inbuilt mechanisms to identify obvious errors. However many datasets have different stages of data entry, for example data are collected and entered locally and then transferred to national datasets. It is clear that the more often the data are handled, the higher the risk of error. Particular issues for data quality are:

- i. Missing data – for example, with birth data, name of the father may be missing, and with marriage data, different religious groups manage information differently and it can take longer to receive information from some religious groups.
- ii. Lack of data – this is a particular issue when datasets rely on self-reporting, for example non-fatal accidents or notification from a clinician (for example communicable diseases)
- iii. Incomplete records – for example an individual on the cancer register may die of another cause.

Confidentiality is an important issue for most agencies. Removing any personal information (for example date of birth, name and address) before releasing the data ensures confidentiality. Some agencies (e.g., Cancer Registry) stated that all information is available to selected people who sign a confidentiality declaration (for example Department of Health officials).

Where the dataset relates to small numbers, for example individuals with a specific disease in a particular location, this could lead to the identification of these individuals. Agencies deal with the issue of small numbers in a variety of ways.

- Data not released – This means that data below a certain cut off point are suppressed. A variety of cut off points are used, for example the Cancer Registry do not release data if the denominator is less than 1000 or there is a count of less than 5 in any area. The CHI Patient Survey do not publish data relating to less than 30 individuals and the Department of Health use a cut off point of 50 for the Prescription Costs Analysis dataset.
- Rounding up data - The Department for Works and Pensions round data to the nearest 100 but have a specialist statistical team who manage small area data, for example by rounding to the nearest 5 at ward level.

- Data published at large geographical area – the British Crime Survey (Home Office) is only published at Police Force Area level to avoid identification of individuals. The Office for National Statistics also publishes data at a regional level (Contraception and Sexual Health dataset and Mental Health of Carers dataset).

These methods of managing the data to avoid identification of individuals can create challenges for assessing the health of rural communities. Rural communities by their nature have small populations and therefore are more likely to be subject to the suppression of data, rounding or being released at large geographical areas. This highlights the importance of Defra having access to individual level data for any analysis of inequalities in rural areas.

6.4.3 Rural relevance of the data

The agencies were asked a series of questions to explore the rural relevance of their datasets. The majority of agencies do not have a specific remit to gather data on rural areas. However in all routine datasets at national level rural areas are covered. Some agencies have the flexibility to explore rural areas in more depth in particular datasets. For example, the Health Protection Agency, which collects the Communicable Diseases dataset, may be triggered to study a rural area if a pocket of disease (e.g. zoonoses) was emerging.

In most cases, although rural data are included in datasets, agencies do not distinguish between urban and rural areas. The majority of the agencies do not have their own definition of rurality although they could attach a national definition if required, probably through postcode filters (ONS Cancer Registry). Agencies stated that they had applied definitions in the past, for example the ONS categorisation of Health Authorities but that organisations and boundaries have changed therefore now it may not be possible to do this.

Just over a third of the respondents were aware of the forthcoming classification of urban and rural areas based on Census Output Areas, However almost half were not aware of the new classification. When asked if plans were in place to analyse the dataset using the new classification six respondents commented that they would apply the new classification if they thought it was advantageous and useful in improving ways of working. This highlights the need for promotion of the new classification. Ten respondents did not have any plans in place to apply the new classification and one agency highlighted the fact that their data are at too high a level for it to be worth applying the new classification.

In general, most data are collected at a large geographical scale. Eight respondents said that Local Authority level is the smallest geographical unit at which data are available, another four stated that data were available at electoral ward level and seven stated that Primary Care Trust was the unit of data collection. Government Office Region (2 respondents), Police Force Area (2 respondents), County District Level (1 respondent), County/Borough

Unit Level (1 respondent), Practice Level (1 respondent) and Secure Unit Level (1 respondent) were also stated.

Respondents were asked their opinion on whether data were representative of rural areas. Responses were positive, though generally only in that the data are collected at national level and therefore cover rural areas. Interestingly the Commission for Health Improvement replied that in the Patient Survey they get better and more complete responses from rural areas.

Over half of the agencies responded negatively to the question over whether urban-rural or intra-rural differences are identified. This is partly because these differences are not looked for in the data unless a specific request is made. The respondents that answered positively to one or both of these questions were: ONS Cancer Registry, Home Office British Crime Survey, Commission for Health Improvement Patient Survey, DoH Prescription Costs Analysis, DoH Personal Social Services Expenditure and DoH Community Care Statistics.

Finally, respondents were asked about the possibility of linking their dataset to other sources of data in order to facilitate an analysis of rural issues. Respondents were largely negative about this, explaining that it would be a difficult task incorporating the potential ethical responsibilities, which would mean obtaining permission. However seven respondents answered positively: ONS Cancer Registry, ODPM Housing Statistics, Defra Agricultural Census, Department of Health Prescription Cost Analysis, Department of Health Health Survey for England, National Primary Care R&D Centre Research Resource Centre, Department of Health Community Care Statistics Home Care Services for Adults.

The discussion with agencies reinforces the findings of the statistical review in that many national datasets are available at a large geographical unit and therefore are unlikely to be sufficiently rurally sensitive. The discussions have also confirmed the challenges that maintaining confidentiality poses in making data available for analysis. A number of additional findings have also emerged which are important for the analysis of rural health issues. Few agencies have an existing remit to examine their data by urban/rural classifications. Encouragingly a small number of agencies stated that they would be willing to do this if they could see the relevance. Less encouraging however, is the fact that there does not appear to be widespread awareness or preparation about applying the new definition of rurality. Consequently there is a need for the promotion of the new definition of rurality and possibly a directive in order to ensure that a rural-urban analysis is undertaken.

6.5 Identification of future research at a small area scale

To identify future sources of data that may provide evidence on rural health issues, information was sought on current and planned research in the UK. Very few current or planned studies were elicited during the search (details of the studies are in Appendix E).

The majority of projects that emerged are at a PCT or regional level. Only one research project is operating at a national level – the Rural Services Customer Review funded by Defra. Consequently there is little information forthcoming at a national level that will supplement the datasets outlined in the inventory.

In terms of subject area, four projects are related to mental health (3 voluntary sector led, 1 health sector led) whilst three projects (all health sector led) are undertaking a general profiling of the health and well-being of their rural communities. Under the NHS Plan, Primary Care Trusts have a key task of ‘improving the health of the community’. As part of this process, PCTs will be gathering information on the health and well-being of their populations to support the development of Health Improvement Programmes. It is probable that the studies identified here are part of this strategic process. It is likely that such studies are on-going in many PCTs across England and it would be worth exploring the collation of these studies to provide a national picture of health and well-being.

Other projects identified are studied rural services (1), emergency care (1) and improving access (1). Overall there appears to be little research at a national level that will provide additional datasets for analysis of rural health issues. This is possibly not surprising as collating data at a small area scale is resource intensive and therefore expensive. To date rural health issues have not been sufficiently prioritised by research funders to allow large-scale analyses, and historically research has been very localised, requiring minimal finance often undertaken by enthusiastic rural health professionals wishing to gather evidence on a particular local issue. In the short to medium term it would appear pragmatic to add value to existing datasets through funding in-depth analysis from the rural perspective applying the new definition of rurality and appropriate geographical mapping system approaches.

7. Conclusions and recommendations

Our brief included five major components and we discuss the main conclusions and recommendations related to these below:

7.1 Compilation of a comprehensive inventory

A comprehensive inventory of databases on poor health and health service use at national, regional and local levels has been drawn up and the annotated master table is included in Appendix B. In the initial trawl 153 datasets were identified. This was subsequently reduced to 120 through the identification of duplication and incompleteness in the initial trawl. Through

grouping datasets this was further reduced to 103 datasets that are included in the annotated master table. The datasets were categorised into one of two major headings – ‘directly health related’ (with 16 subgroups) and ‘wider determinants of health’ (with 4 subgroups).

To add value to the inventory in terms of the rural perspective, the table has been annotated to provide detail on the type of data collected, its available spatial resolution and the potential that each dataset has in relation to implementing the new definition of rurality. The purpose of the inventory is to provide a detailed overview of datasets that may be relevant to a specific analysis of a rural health issue. The next step will be to use the inventory in addressing the rural health agenda outlined in section 3.

7.2 Planned research at a small area scale

The second objective of this study was to identify any planned research into rural health issues at a small area scale, which would yield datasets for future analysis. Using the rural networks at IRH a widespread call for information on current and planned research was made (results are shown in Table E). There is limited current or planned research that will elicit data at a national level. Defra’s own Rural Services Customer Review was the only project found to be operating at a national level. However the data being gathered are not directly health related. A number of health sector organisations are undertaking surveys of the health and well-being of the populations in their area. To support the implementation of health policy at PCT level, many PCTs are undertaking such health and well-being assessments, highlighting the opportunity to collate information at a national level

7.3 Assessment of datasets for use in a rural analysis and identification of problems

Objectives 3 and 4 of this study required an assessment of the suitability of datasets from a rural perspective and the identification of any problems in using the data to describe health levels in rural areas. This was done as a statistical review undertaken by the Centre for Medical Statistics (GM and NG).

Most of the datasets included in the inventory are ecological in nature. In the methodological appendix (Appendix A) we have drawn attention to two types of logical difficulty of interpretation, which can arise in spatial studies. The *Aetiological Fallacy* arises principally in explanatory studies because Hill’s Criteria for Causality cannot be applied to spatial units. This rather colours our view of the usefulness of the ecological data in the master table if our purpose is to mount a study of disease causation - perhaps an incidence study of some type of cancer. Of course this problem of aetiological interpretation is not peculiar to rural health studies but attends all spatial studies. This leads us to conclude that for explanatory studies the acquisition of anonymised, individual-level data is a priority, for example through improving access to primary care morbidity data. Our findings in relation to the special impact of

the range of anti-disclosive measures, due to confidentiality issues, currently deployed on rural health studies reinforce this conclusion.

The other problem of interpretation is the modifiable area and to counter this we should prefer to have ecological data at a level of spatial resolution, which facilitates re-definition of the spatial units in testable ways, thereby enabling rural health researchers to protect conclusions against the modifiable aerial unit problem. Broadly speaking, the spatial resolution of many of the health datasets contained in the inventory is too crude to inform the rural health agenda and to facilitate the necessary protective checks. This is particularly true of data collected and published at PCT/NHS Trust or local authority/unitary authority (LA/UA) levels. We conclude that rural health researchers will need to focus on accessing and aggregating the underlying microdata and observing the consequent confidentiality constraints on publication. Of course, some ecological datasets do satisfy the prescribed methodological criteria in terms of spatial resolution and rural health researchers can make use of them for *descriptive* studies.

Given the foregoing, it is not possible to give a blanket stamp of approval to individual datasets, rather the usefulness of any dataset will depend on the particular rural health problem to be addressed and the particular hypotheses of interest. If a hypothesis has high aetiological content it is unlikely that the ecological datasets described here will be of interest. In this context we conclude that the investment in the UK's Neighbourhood Statistics programme (Neighbourhood Statistics, W31) is unlikely to produce a return in terms of aetiology (in medical or other types of studies).

For the reasons discussed elsewhere the new definition of rurality is an advance on previous definitions. The discussions with agencies suggest that there is low awareness of the forthcoming definition, but on discussion, there is interest in its implementation as long as *utility* is evident. Our sample survey results suggest the prospective introduction of the new definition will be accepted, but we have no detailed information on the acceptability of its *retrospective* implementation. We conclude that this issue is an area for further research.

7.4 Recommendations for improving data collection and analysis

Seven clear recommendations for improving data collection and analysis are evident from this study:

- i) This report has highlighted the distinction between ecological and aetiological data and their respective ability to answer questions of causality. Whilst not all questions on the rural health agenda seek to explore causality, it nevertheless, remains an essential element of understanding the impact of rurality on health. This distinction needs to be understood by those who commission *and* undertake data collection and analysis and should feature in any inter-agency discussions about the development of datasets to address the rural

health agenda. Wide dissemination of this report and/or a series of briefing seminars should assist in this regard.

- ii) The datasets included in the inventory are largely ecological in nature, however micro-data exist beneath these datasets, which highlights the need to forge confidentiality agreements with the major data providers, on a systematic and on-going basis. In this regard, a national micro data panel for health statistics would be a valuable resource.
- iii) For a descriptive analysis of rural health issues, there is a wealth of datasets included in the inventory, which will be valuable. Most of these cover rural areas as part of a routine national data collection. However researchers need to explore in greater detail the actual sample size for rural areas, as this may not be sufficiently large for analysis. We recommend further statistical review of these datasets.
- iv) The new definition of rurality will be an important tool in addressing the rural health agenda, however Defra will need to champion its cause in other Government Departments, provide incentives for its implementation by external Public Bodies and popularise it among the rural health research community. Whilst the statistical review has shown that applying the new definition of rurality prospectively to many existing datasets is feasible, it is unknown how straightforward a retrospective analysis will be. In rural areas where numbers tend to be smaller, researchers often rely on trend data or calculating 'rolling-averages' over 3 or more years. Consequently more research into the practicalities of retrospective analysis with the new definition is necessary.
- v) The inventory of datasets (presented in the annotated master table) will be a valuable resource to anyone in policy, service or academic settings who wish to consider rural health issues. We would recommend widespread access to this inventory, either over the Internet or by CD ROM. There will also be a need to update this inventory as new datasets emerge or existing datasets are improved, and consideration will need to be given to a programme of work to achieve this.
- vi) This study has provided a valuable platform for future research into rural health by Defra and the wider community interested in rural health issues. More work is now needed to prioritise the research questions that can be answered through the existing datasets and to undertake a research programme that will provide a descriptive analysis of rural health issues. The use of geographical information systems should be considered with strong links to medical statistical modelling expertise. Such a partnership would provide a strong platform from which to address primary research on the rural health agenda.

- vii) In relation to rural health studies of causality, further technical statistical research in this area is required to evaluate fully the role of micro-data and the limitations of small area statistical data on the types of relevant aetiological studies, which are required to address the rural health agenda.

References

Acheson, D. (1998) *Independent inquiry into inequalities in health report*. The Stationery Office, London.

Asthana, S., Brigham, P., Gibson, A., (2002) *Health resource allocation in England: what case can be made for rurality?* University of Plymouth, Plymouth.

Besag, J., et al (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 43:1-59

Bohning, D., (1999) *Computer Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease mapping and others*. Chapman & Hall. Boca Raton

Buchan, T., (2002) *A cultural and spatial analysis of adolescent substance misuse in rural Wales: final report*. Institute of Rural Health, Newtown, Wales.

Campbell, N., et al (2000) Rural factors and survival from cancer: analysis of Scottish cancer registrations. *British Journal of Cancer*. 82 (11), 1863-1866

Chapman, P., et al (1998) *Poverty and exclusion in rural Britain: the dynamics of low income and employment*. Joseph Rowntree Foundation, York.

Charles, et al (2001) *Feasibility Study on the Availability and Associated Issues of Small Area Health and Socio-Economic Data. Welsh SHARPS initiative*. 1-84

Clayton, D., Kaldor, J., (1987) Empirical Bayes estimates of age-standardised relative risks for use in disease mapping. *Biometric*; 43:671-681

Clayton, D., (1991) A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* 47: 467-485

Clayton, D., Berardinelli, L., (1992). *Bayesian Methods for Mapping Disease. In Geographical and Environmental Epidemiology: Method for Small Area Statistics*. OUP,

Cloke, P., et al. (1997) *Rural Wales: community and marginalisation*. University of Wales Press, Cardiff.

Cox, D.R., (1970) *The Analysis of Binary Data*.. Methuen; London.

- Cox, D.R., (1972) Regression models and life tables (with Discussion) *JRSS B*. 34: 187-220
- Cressie, (1993) *NAC Statistics for Spatial Data* . Revised Edition. Wiley; New York.
- Goldstein, H., (1995) *Multilevel Statistical Models*: Edward Arnold: London
- Daly, et al (1986). *Childhood Leukaemia in the Republic of Ireland Mortality and Incidence*. RPI -DOH Report. 3-90
- Deaville, J., (2001) *The nature of rural general practice in the UK – preliminary research*. A joint report from the Institute of Rural Health and the General Practitioners Committee of the BMA.
- Diggle, P.J., (1990) A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a pre-specified point. *JRSS Series A*.153:349-362.
- Diggle, et al (1998) Model-based Geostatistics. *JRSS Series C* 47: 299-350.
- Dunn, J., et al (1998) *Developing indicators of rural disadvantage*. Rural Development Commission, Salisbury. Research Report Number 36
- Evans, AS. (1978) Causation & disease. *The American Journal of Epidemiology* 108, 4: 249-
- Gardiner, (1983). *Atlas of Cancer Mortality in England and Wales* Wiley New York
- Gelman et al (1995). *Bayesian Data Analysis*. . Chapman & Hall. London
- Gillon, N., MacKenzie, G., (Embargoed until publication, expected Summer 2004) *The North Staffordshire Breast Cancer Study: Consultation Report*, Centre for Medical Statistics, Consultation Report. October 2003
- Goldstein, H., (1995) *Multilevel Statistical Models*. Arnold London
- Green, B. M. R., Miles, J. C. H., Bradley, E. J., Rees, D. M. (2002) *Radon Atlas of England and Wales*. Chilton, NRPB-W26
- Greig, M., Pemberton, J., Hay, I., MacKenzie, G., (1980) A prospective study of the development of coronary heart disease in a group of 1202 middle-aged men. *J. of Epidem. and Comm. Health*. 34: 23-30
- Hawton, K., et al (1998) *Suicide and stress in farmers*. The Stationery Office, London
- Healthy Norfolk 2000 (1998) *Rural health action area pilot project – phase 1. Needs assessment*. A report for Health Norfolk 2000

Higgs, G., (1999) Investigating trends in rural health outcomes: a research agenda, *GEOFORUM* 30 (3): 203-221 AUG 1999

Hill, A.B., (1965) The environment and disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. 58:295-2300.

Holt, D., Steel, D. G., Tranmer, M., (1996) "Area Homogeneity and the Modifiable Aerial Unit Problem". *Geographical Systems* . Vol 3: 181-200.

Howard, G., (2003) *Arsenic, Drinking Water and Health Risk. Substitution in Arsenic Mitigation: a Discussion Paper*. WHO Geneva.

Jones, A., Bentham, G., (1992) Emergency medical service accessibility and outcome from road traffic accidents. *Public Health*, 109, 169-177.

Jones, A., Bentham, G., (1997) Health service accessibility and deaths from asthma in 401 local authority districts in England and Wales 1988-1992. *Thorax*, 52(3), 218-22.

Kalbfleisch, J.D., Prentice, R., (1980) *The Statistical Analysis of Failure Time Data*. Wiley, New York

Lawson, A.B., (1993) On the mortality events around a pre-specified fixed point. *JRSS Series A*. 156:363-377

Lawson, A.B., Williams, F.L.R., (1994) Armadale: a case study in environmental epidemiology. *JRSS Series A*. 157

Lawson, Cressie, (1999) Spatial Statistical Models for Environmental Epidemiology. In *Handbook of Statistics: Bio-Environmental and Public Health Statistics*. North Holland

Machin & D'Agostina. Ed. Statistics in Medicine. (2000) Disease Mapping with a Focus on Evaluation. Vol 19, Nos 17/18:2201-2593

Mackenzie, G., (1989) The Bradford Hill Legacy, *The Ulster Medical Journal*. 67, 1: 66-67

Mackenzie, G., (1996) Regression models for survival data: the generalised time dependent logistic family. *JRSS. Series*, 45:21-34

Openshaw, S., (1997) A geographical solution to scale and aggregation problems in regio-building, partitioning and spatial modelling. *Transaction of the Institute of British Geographers*. NS 2: 459-472

Openshaw, S. Rao, L., (1995) Algorithms for re-engineering 1991 Census geography. *Environment and Planning A* 27: 425-446.

PODAR Committee (1989) *Investigation into patterns of disease with possible association with radiation in Northern Ireland*. Belfast HMSO 1-146

Pooling Project Team (1989) Relationship between blood pressure, serum cholesterol smoking habit etc. to incidence of major coronary events: final report of the pooling project. *Journal of Chronic Disease* 31, 4; 201-306.

Rawles, J., et al (1998) Call to needle times after acute myocardial infarction in urban and rural areas in northeast Scotland: prospective observational study. *BMJ*, 317: 576-8

Robinson, (1950) Ecological correlation and the behaviour of individuals. *American Sociological Review*. 15 :351-357

Shucksmith, M., (1999) *Poverty and social exclusion in rural Britain*. Section I in The Rural Group of Labour MPs (1999) *Rural audit: a health check on rural Britain*. The Rural Media Company, Hereford

Starke et al (1997) The effects of access factors on breast screening attendance on two Scottish islands. *Health Bulletin* 55(5): 316-21

Woollett, S., (1990) *Counting the rural cost – the case for a rural premium*. NCVO Publication

Key website references

Table:		Website References
Key	Topic	URL
W1	Office for National Statistics	www.statistics.gov.uk
W2	UK Association of Cancer Registries	www.ukacr.org.uk
W3	National Cancer Plan	www.doh.gov.uk/cancer/cancerplan.htm
W4	Trauma Audit & Research Network (TARN)	www.tarn.ac.uk/main.asp
W5	Vital Statistics	www.statistics.gov.uk/statbase/Product.asp?vlnk=539
W6	ONS Geography - Beginner's Guide	www.statistics.gov.uk/geography/beginners_guide.asp
W7	ONS Urban/Rural Best Practice Guidelines	www.statistics.gov.uk/geography/best_practice.asp
W8	ODPM Geography - Project report	www.statistics.gov.uk/geography/geog_initiatives.asp
W9	ONS Census Geography	www.statistics.gov.uk/geography/downloads/geoadshowpaper.pdf
W10	The definition of urban areas for the 2001 Census	www.statistics.gov.uk/census2001/pdfs/urban_area_defn.pdf

W11	Best Practice Advice	www.statistics.gov.uk/geography/downloads/User%20Guide%20_27AugONS.pdf
W12	Defra	www.defra.gov.uk/esg/work_htm/publications/cs/ruralinfo_web/General/Outline.pdf
W13	Oxford -CA	www.odpm.gov.uk/stellent/groups/odpm_planning/documents/page/odpm_plan_609188-03.hcsp
W14	Data Protection - Information Commissioner	www.informationcommissioner.gov.uk/
W15	Caldicott Report	www.doh.gov.uk/ipu/confiden/
W16	Census Office	www.nisra.gov.uk/census/pdf/ag0203.pdf
W17	ONS Microdata Panel	www.statistics.gov.uk/about/NS_ONS/ONS_microdata_releases.asp
W18	National Statistics Code of Practice	www.statistics.gov.uk/about_ns/cop/default.asp
W19	Protocol on Data Access and Confidentiality	www.statistics.gov.uk/about/consultations/general_consultations/downloads/Protocol_on_Data_Access_Confidentiality.pdf
W20	LREC. Local Research Ethical Committee	www.corec.org.uk/LRECContacts.htm
W21	MREC Multi-centre Research Ethical Committee	www.corec.org.uk/howToApply.htm
W22	NHS Current code of Practice on Confidentiality	www.doh.gov.uk/ipu/confiden/protect/index.htm
W23	Irish Statistics Act 1993	www.irishstatutebook.ie/ZZA21Y1993.html
W24	ELSA	www.natcen.ac.uk/elsa/
W25	RCGP - Primary Care Research Networks	www.rcgp.org.uk/research/pcrta/exec_summ.asp
W26	Clinical Governance	www.modern.nhs.uk/scripts/default.asp?site_id=18&id=3962
W27	Research Governance framework for England	www.doh.gov.uk/research/rd3/nhsrandd/researchgovernance.htm
W28	Black Report Health Inequalities 1980	www.sochealth.co.uk/history/black.htm
W29	Acheson Inquiry Health Inequalities 1997	www.doh.gov.uk/ih/ih.htm
W30	DOH Tackling Health Inequalities	www.doh.gov.uk/healthinequalities/programmeformation/index.htm
W31	Neighbourhood Statistics Programme	http://neighbourhood.statistics.gov.uk/default.asp?nsid=false&CE=True&SE=True

Background statistical considerations

Risk factor evaluation in rural health studies

In order to understand how these logical issues impact on the use of existing databases and on the evaluation of risk factors in particular rural health problems we contrast the analysis of epidemiological and ecological studies in more detail.

1. Aetiological Epidemiological Studies

From a classical Epidemiological standpoint, the assessment of a putative causal hypothesis is carried out in a rigorous framework outlined, elegantly, by Sir Austin Bradford Hill (1965). Hill set forth a set of criteria, his eight Criteria for Causality, for assessing the *aetiological significance* of a putative risk factor for a particular disease (Table 1).

These criteria are based on the Henle-Koch criteria, synthesised by Evans (1978) as an integrated set of criteria for causation applicable to communicable and non-communicable disease. They are guideline criteria, which a putative risk factor must satisfy if it is to be regarded as a potential cause of the disease under study. Such criteria are required to distinguish spurious associations from aetiological relationships.

Table 1: Summary of Bradford Hill's Criteria for 'Causality'
1. Strength
2. Consistency
3. Specificity
4. Temporal Sequence
5. Biological Gradient
6. Plausibility/Coherence
7. Experimental Evidence
8. Analogy
9. <i>Independence of Effects</i>

The classical longitudinal epidemiological study starts with disease free patients, measures the putative risk factors, X, at baseline and observes the development of the disease outcome, Y, over time. In formal statistical terms we model the disease process, Y, *given* our knowledge of the risk factors, X, measured at baseline i.e. $Y|X$ (the vertical bar mean 'given'). In this way, we note, that Hill's 4th criterion, that the putative cause, X, must precede the effect, Y, is automatically satisfied. In general, the components of X - the risk

factors - are chosen, when possible, with respect to criteria 5-8. The quantitative assessment of the influence of the risk factors in any study is usually conducted using criteria 1-3 in the framework of a statistical model invoked to estimate the influence of the components of X singly *and* in combination. However, in this framework we must add a new criterion to Hill's original eight, namely, the *independence of effects* in a statistical model (MacKenzie, 1996). It is this new criterion that enables us to determine whether an association between the outcome and a risk factor is real or whether it can be explained by the influence of other risk factors. We may argue that such a criterion is an essential tool when working in a framework of multiple association.

Thus, for example, in relation to the incidence of coronary heart disease, CHD, (Pooling Project, 1989, Greig *et al* 1980) we may hypothesise that X comprises two factors, diastolic blood pressure (x_1) and total tobacco exposure (x_2), i.e. $X = (x_1 \ x_2)$.

Then, if we fix the follow-up period (e.g., 5 years) our statistical model for disease incidence is the standard multiple linear logistic model (Cox, 1970) in which the logarithm of the Odds on developing the disease is modelled as:

$$\log(\text{Odds on developing CHD}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

The Odds is simply probability of developing CHD divided by the probability of not developing CHD, during the study. The magnitude of the β_1 and β_2 coefficients measure the strength of the influence of the factors on the log (Odds) function. A large positive value of a coefficient indicates that higher values of the factor increase the probability of CHD. A large negative value of a coefficient shows that as the factor increases the probability of CHD is decreased, i.e., the factor is protective. A zero value for a coefficient indicates that the corresponding factor does not influence the probability of CHD. The coefficient β_0 is an intercept term.

Thus, one goal of the statistical analysis is to estimate the values of the β coefficients (the effects) and to eliminate redundant factors, i.e. those in which the corresponding β 's are close to zero. The statistical analysis provides numerical estimates of the β 's and a criterion (standard error) with which to test whether they are significantly different from zero.

For 2 risk factors there are four possible models of outcome, namely:

$$\log(\text{Odds}) = \beta_0 \tag{1}$$

$$\log(\text{Odds}) = \beta_0 + \beta_1 x_1 \tag{2}$$

$$\log(\text{Odds}) = \beta_0 + \beta_2 x_2 \tag{3}$$

$$\log(\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{4}$$

(bold indicates effect is non-zero)

Model (1) asserts that the probability of developing CHD does not depend on either factor i.e. it is a constant value; model (2) that it depends only on

diastolic blood pressure; model (3) that it depends only on total tobacco exposure; and model (4) that it depends on both factors.

It is model (3) which allows us to implement the 9th criterion. In particular, if β_1 and β_2 are (significantly) non-zero, both coefficients are required. Then β_1 is called the independent effect of diastolic blood pressure, because it represents the "independent" influence of diastolic blood pressure when total tobacco exposure is in the model. Similarly, β_2 is called the independent influence of total tobacco exposure, when diastolic blood pressure is in the model. In this case we know that both factors are required, i.e. the influence of one cannot be explained by the other and *vice versa*.

Thus the aetiological significance of a factor rises if its influence is undiminished when the effects of other factors are taken into account in the model.

Suppose we extend the analysis by adding an Urban/Rural split ($x_3=1$ for Urban and $=0$ for Rural) to the set of putative risk factors, hypothesising that the incidence of CHD is higher among those individuals living in Urban areas. Suppose, in addition, we fit the following sequence of models

$$\log (\text{Odds}) = \beta_0 + \beta_3 x_3 \quad (5)$$

$$\log (\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_3 x_3 \quad (6)$$

$$\log (\text{Odds}) = \beta_0 + \beta_2 x_2 + \beta_3 x_3 \quad (7)$$

$$\log (\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (8)$$

Then we may deduce from model (5) that there is significant Urban/Rural influence which survives the inclusion of the known risk factor diastolic blood pressure [model (6)] and separately survives the inclusion of total tobacco exposure [model (7)], but not, however, the simultaneous inclusion of these two known risk factors, since β_3 in model (8) is no longer significantly different from zero. Thus we may conclude that the apparent effect of Urban living on the development of CHD is wholly explained by the joint influence of diastolic blood pressure and total tobacco exposure. The joint influence of these two factors is said to *abolish* the influence of the geographical factor.

Notice that our conclusion would have been quite different had we stopped at an earlier model, since then we would have reported the presence of a spurious Urban/Rural effect. In this latter case we should have failed to account properly for the influence of potentially confounding risk factors.

The effect of controlling for potentially confounding risk factors may not be to abolish the influence of the factor of interest, but rather to reduce the magnitude of its effect. Thus the complete elimination, or reversal (in sign), of an effect are merely extreme examples of effect modification.

This illustrates some of the principles involved in risk factor evaluation, presages the complexity, which can arise when the number of potential risk factors is large and highlights the logical pitfalls involved. The ecological fallacy is also related to a failure to properly account for confounding variables, but in a rather different sense.

There are of course many other types of outcome other than disease incidence, e.g. survival (Cox, 1972, Kalbfleisch & Prentice, 1980; MacKenzie 1996), which are relevant and there are a wide variety of statistical models available to handle these.

2. Ecological Studies

Unfortunately the analysis and interpretation of such ecological data are not straightforward, since relationships at an aggregated level do not necessarily hold at the level of the individual. Individual-level procedures, such as controlling for confounding risk factors and effect modification, do not necessarily lead to the same results when applied to spatial units as when applied to individuals. This is due to the fact that area variation in individual outcomes and risk factor values are ignored in ecological studies.

It is this within-area variation that may generate the "Ecological Fallacy" whereby an association found at the ecological level does not hold for individuals. (Robinson, 1950).

We distinguish this problem from the well-known Modifiable Areal Unit Problem (MAUP) which arises in ecological studies when the associations found are not invariant to the choice of spatial unit (e.g., number and boundaries). Thus the MAUP arises when two ecological studies yield different results due to the choice of different spatial units. (Holt, 1996).

It is now apparent that in contrast with aetiological studies, the probability of each individual developing the disease cannot be evaluated in an ecological study.

From this perspective, ecological studies clearly lie outwith the logical framework in which Hill's criteria for 'causality' were developed and cannot therefore be regarded as aetiological in nature. Rather, they should be viewed primarily as *descriptive* and *exploratory* tools revealing spatial patterns and perhaps suggesting further aetiological hypotheses for analysis.

In health studies based on ecological data we shall refer to the "Ecological Fallacy" as the "Aetiological Fallacy" in view of the aetiological constraints noted above.

3. Disease Mapping & Exploratory Studies

A major use of ecological studies is to describe the spatial patterns of disease incidence and/or mortality, i.e. to answer the "Where" question. The use of shaded or colour-coded medical maps (Choropleths) has become an indispensable tool for visualising the spatial distribution of disease in the community.

An early example of their use in Epidemiology was the Atlas of Cancer Mortality in England & Wales (Gardiner *et al*, 1983) which presented the

spatial distribution of Standardised Mortality Ratios (SMRs), by Local Authority (LA) Area, for a comprehensive set of cancers, based on deaths over the 11 year period 1968-78. The Atlas presented mortality maps by persons and for each sex separately at LA and County levels.

Daly (1986) mapped the mortality and incidence of childhood leukaemia in the Republic of Ireland by variety administrative geographies (District Electoral Divisions (DEDs) and Urban and Rural Districts). With co-workers he described patterns observed along the Eastern coastal strip from Carlingford Lough in the North East to Mizen Head in the South West.

MacKenzie (PODAR, 1989) mapped the incidence and mortality of leukaemia in Northern Ireland in two periods 1968-76 and 1977-85, presenting the spatial distributions of crude and age-sex adjusted rates by electoral wards, for various sub-types of leukaemia. The null hypothesis of no difference in the adjusted incidence (and mortality) rates between coastal and non-coastal wards was tested and found to be consonant with the data in both periods. The test required individual level data (age, sex and electoral ward) on all cases of leukaemia and similar data for the population at risk. This *intrinsically* geographical hypothesis was predicated on the idea that the incidence in coastal wards might have been elevated as a result of radioactive discharges from the Sellafield nuclear reprocessing plant in Cumbria.

It is this type of, *a priori*, spatial hypothesis that the ecological study is designed to address. Note however, that in the PODAR study, the aetiological content of the hypothesis is low. Had the null hypothesis, of no difference, been rejected it would not necessarily have implied any association with radioactive discharges. Rather, it would have implied that the incidence and/or mortality rates were higher in coastal wards and, of course, many competing explanations exist. However, it is likely that such an ecological finding would, *at that time*, have led to further aetiological-based research in the form of retrospective case-control studies involving individual-level data.

The statistical analysis of ecological data is similar to that of aetiological data with many formal structural elements in common. Typically, an appropriate statistical model for the response Y (mortality or incidence rates) in each spatial unit is formulated in terms of context related X variables (e.g. proportion of males, proportion of persons over age 60 years, Urban/Rural split average level of deprivation, etc). However, there are also some key differences. Whereas, in an aetiological study of non-communicable disease the response of any individual (diseased/not diseased) is independent of the responses of the other individuals, in an ecological study neighbouring spatial units often have similar values of the response variable. This spatial correlation must be modelled explicitly in any statistical analysis (Cressie 1993, Diggle 1990, Diggle *et al* 1998). If the ecological study is based on units with low spatial resolution the correlation structure may be described more appropriately by a hierarchical or multi-level model structure (Goldstein, 1995).

During the eighties interest in Medical Mapping, Disease Mapping, Medical and / or Epidemiological Geography grew apace. This was fuelled, in part, by the rise in computer power and the increasing availability of mapping software

and Geographical Information Systems (e.g. MapMaster, Map-Info and Arc-Info). Statisticians became interested in a variety of technical issues related to modelling, spatial imaging, smoothing, covariance structure and representation.

Clayton and Kaldor (1987) published a seminal paper on empirical Bayes estimates of age-standardised relative risks for use in disease mapping. Thus was the connection between Bayesian Modelling and Medical Mapping made. This area, driven by advances in Bayesian computation, based on Monte Carlo Markov Chain (MCMC) methods, is now a virtual industry (Clayton 1991, Besag 1991, Clayton. *et al*, 1992; Lawson, 1993, Lawson 1994, Böhning 1999, Gelman *et al* 1995, Lawson & Creesie 1999). A valuable reference is the special issue of Statistics in Medicine (2000), which is entirely devoted to disease mapping.

These technical advances allow richer classes of statistical models (e.g. space-time models and age-period-cohort models) to be deployed. They also facilitate the extraction of the maximum amount of information available in the ecological data and permit sensitivity of conclusions to model assumptions to be checked.

However, they cannot eliminate the fundamental *logical* difficulties associated with the interpretation of conclusions based on ecological data.

4. The Nature of Spatial Risk Factors

The foregoing material illustrates some general problems attending the evaluation of spatial risk factors and highlights the difficulties involved. Clearly in aetiological studies the problems are less. Then we have access to individual-level data including appropriate spatial information (post-code or geo-code) which can enter the analysis on the same basis as other risk factors and be evaluated in an appropriate statistical model. If the effect of the spatial factor is not explained by competing risk factors we may conclude that the spatial factor has aetiological significance.

When we are studying the incidence (or survival) of a particular disease we shall be interested in the exact nature of the spatial effect. If the spatial factor is simply the Urban/Rural split we shall have learned that there is a difference in disease incidence, between urban and rural individuals, which is not explained by the other factors in the statistical model. The next obvious question is what underlying factor has caused this effect? The implication being that there may be nothing *intrinsically* spatial, *per se*, in the aetiology of the condition.

For example, Gillon & MacKenzie (2003) in their study of breast cancer survival in the West Midlands showed that a battery of nine clinical and biological factors explained the differences in five-year survival in all but two of the 13 local health authorities studied. Survival in these two LHAs was significantly better but the reasons for this are not thought to be related to the geography of the LHAs. Rather, the explanation is more likely to lie in the

organisation of services and/or clinical care or in unmeasured risk factors in the patients studied.

Similarly with ecological data the detection of a spatial pattern of disease incidence may merely reflect artefact - contamination from a reprocessing plant, leakage from a disused mining facility or industrial pollution from a large manufacturing facility. The detection of these types of effects requires the damage to be sustained in time and its geographical distribution to be more widespread than the spatial resolution of the ecological study.

In general, in health studies, it may not be particularly easy to conceive of a *natural* spatial effect, which has specific aetiological significance. Radon in rock (Green et al, 2002) and arsenic in well water (Howard, 2003) are two obvious examples of *naturally* occurring spatial factors, which adversely affect health.

Accordingly spatial effects often may be viewed as surrogates for the effects of other, unmeasured, risk factors and this observation lends further justification for adding the 9th criterion (Table 1).

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
DIRECTLY HEALTH-RELATED				
<i>Vital Statistics</i>				
Birth Statistics: Compiled by ONS - annual live birth counts in Table VS4 available from Vital Statistics Output Branch (VSOB); Live birth counts also by sex, for 1998, available on Neighbourhood Statistics website. ONS volume FM1 gives numbers of stillbirths, fertility rates, birth weights (live) and conception numbers/rates at HRO and metropolitan county levels for areas of usual residence.	Annual statistical collections of all registered births	Ward	✓	No current rural/urban breakdown. However new definition could be readily applied to ward level data. Bespoke vital statistics outputs are also available at request to VSOB at ONS if confidentiality is not compromised.
Mortality Statistics: Compiled by ONS - annual counts of deaths (except stillbirths and infant deaths under 28 days) in Table VS4 available from Vital Statistics Output Branch (VSOB). Similar death counts, but also by sex, available on Neighbourhood Statistics website for 1998. Deaths by selected cause available from VSOB in Table VS4D.	Annual statistical collections of all registered deaths	Ward	✓	No current rural/urban breakdown. However new definition could be readily applied to ward level data. Bespoke vital statistics outputs are also available at request to VSOB at ONS if confidentiality is not compromised.
Mortality Statistics: Compiled by ONS - numbers of deaths by cause using ICD-10 chapters, age and sex available from VSOB in Table VS3. Infant mortality figures available from VSOB in table VS5. Death numbers and rates by age and sex, standardised mortality ratios and place of death given at HRO, GOR and metropolitan & non-metropolitan county level in ONS volume DH1.	Annual statistical collections of all registered deaths	UA/CD	☑	No current rural/urban breakdown. This mortality data only available publicly down to unitary authority/county district level. However bespoke vital statistics outputs may be available with rural indicators added at smaller areas, depending on confidentiality constraints. Data could not be published at smaller areas.
Linked Births & Infant Deaths: Compiled by ONS. An infant death is defined as death under 12 months. Data covers information on birth and death certificates including age and sex. Socio-economic variables using postcode information and mother and father's occupations (if known and given on certificates) can be derived by ONS.	Linkage of infant deaths (less than 12 months) in England & Wales to corresponding birth records by ONS, which has been undertaken since 1975	Ward if not disclosive	✓	No current rural/urban breakdown. Such a breakdown could be used if groups defined were not disclosive and postcode information on birth and death certificates could be used to attach rural/urban indicators to the dataset. Data can be released potentially at ward level if this was not disclosive, however all requests for data are considered on a case by case basis by the Microdata Release Panel within ONS, before data are released outside of the organisation, and will depend on the research purposes and ethical/confidentiality issues.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Marriage statistics: Compiled by ONS in annual FM2 volume. Includes: numbers of marriages; previous marital status of bride and bridegroom by age; type of marriage ceremony. Divorce statistics are also included in same volume but only at national level.	Annual statistical collections of all marriages registered	UA/CD	<input type="checkbox"/>	No current rural/urban breakdown. Data for marriage is available publicly from ONS down to county (registration) district and unitary authority levels. Data does not contain any residence details, so data cannot be broken down into smaller areas. Divorce statistics are only given at national level as the only location detail held is for the court granting the divorce.
Population & Internal Migration				
National Census and Neighbourhood statistics: 2001 census carried out by ONS, based on usual resident population contains detailed population including self-reported general health over last 12 months, limiting long-term illness, health and provision of unpaid care. Data available from Neighbourhood Statistics part of ONS website. Also contains information on age, marital status, ethnic group, access to transport, religion, qualifications, accommodation and amenities, household composition and employment details. However data soon outdated as census only decennial. Neighbourhood statistics data, deposited by other government organisation, on ONS website can be linked to census data (mostly at ward level) and includes: benefit claimants & employment, education, some hospital episode statistics, housing, deprivation indices.	Decennial Census returns – most recent 2001	Output area	<input checked="" type="checkbox"/>	Currently, the Countryside Agency (CA) ward level definition of rural areas is included in the Neighbourhood Statistics section of website. New rural/urban definition will be available at Output Area level, and so can be readily attached to census data. The census output areas are the smallest areas from which counts can be taken from census data, and are limited to a minimum of 100 persons and 40 households. Blurring of the data, including small cell count adjustment, is carried out by ONS to prevent disclosure of information. It may be possible for bona fide researchers to enter contracts with ONS Census Office to receive un-blurred data for analysis, but results could only be published above normal limits.
GP Relevant Populations: Data collected by DH in April 2002 on both GP relevant populations & residence relevant populations for Strategic Health Authorities & Primary Care Trusts at 2002 boundaries. Numbers reconciled to ONS mid-2001 population estimates.	Data extracted from Family Health Services (FHS) registers	PCT	<input type="checkbox"/>	No rural/urban breakdown used. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Internal Migration estimates: Estimates, since 1975, by ONS of internal migration in England & Wales using NHS registers. Data consists of numbers of inflow/outflow and net balance into each area by age group and sex. Data represents a snapshot taken on 31 st July each year and compared with snapshot data from year before, which is then matched by NHS number. Any record with postcodes that are different between years is extracted as a migrant move.	Data extracted from the NHS Central Register.	UA/CD	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new classification. Individual patient data cannot be released by ONS – DH releases the data to them solely for production of population statistics. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Comprehensive General Practice/PCT Databases				
<p>National database for Primary Care Groups and Trusts: Practice level data: Taken from General Medical Services (GMS) data up to 2001 – gives GP characteristics, practice activity, patient characteristics. Also links ward level census variables to practice data. PCG/T level data: clinical indicators (2000); various measures of deprivation; PCG resource allocations (up to 2001/02). Database access is currently only available to NHS staff and the academic community who register.</p>	Database based at University of Manchester using data from various sources including from ONS and DH	Practice	✓	GMS data defines a rural patient if living more than 3 miles by usual route from their GP's main surgery - used for allocating extra payments to GPs with at least 20% of 'rural' patients on list to cover extra expenses. Using this definition, it is possible that urban GPs can have 'rural' patients, so not useful for geographically based urban/rural breakdown. Database managers are interested in using new rural/urban definition if found to be useful - could be attached using practice postcode.
<p>General Practice Research Database (GPRD): Database held by the MHRA. Individual patient data collected since 1987 which includes: demographics of patient; medical diagnosis and comments; prescriptions; events leading to drug/treatment withdrawal; hospital records; treatment outcomes; other patient care information – e.g. height, weight, immunisation, smoking status and lab results. Data are available online – prices detailed on website.</p>	Database of data obtained from UK general practice computer records using Vision software – approx 3 million patients (5% UK population)	Areas with 3 or more practices	<input checked="" type="checkbox"/>	No current breakdown of rural and urban areas, and no current plans to use such a breakdown as standard within the database. Rural/urban indicators may be added to practice postcodes (at lowest level) and such indicators have been used in the past. Ad hoc analyses of the data can be carried out by the GPRD research team at a charge. Any studies using GPRD data for publication must be approved by the Scientific and Ethical Advisory Group (SEAG).
<p>The Health Improvement Network (THIN): Database resulting from collaboration between the two companies EPIC and In Practice Systems. Contains detailed individual patient data. Data are collected continuously and relate to early 1990s to 2003. Data are available from EPIC as flat file ASCII datasets.</p>	Database of Vision software patient data from participating practices – presently 195 practices - about 1.5 million patients	Areas with 3 or more practices	<input checked="" type="checkbox"/>	No current rural/urban breakdown. Rural/urban indicators may be added to practice postcode if anonymity is maintained in breakdown. Such ad hoc analysis can be carried out by the THIN research group. Patients are anonymised and the linking code cannot be broken without the consent of the patient and the co-operation of the GP, so patient level linkage of rural/urban indicators is not practical.
NHS Performance/Waiting Times				
<p>Ambulance Services: Data collected by DH and includes: numbers of patient ambulance journeys; numbers of emergency calls; numbers of emergency calls resulting in response call-out; response times by ambulance; arrival times in urgent journeys; arrival times in relation to requested arrival time; arrival time no later than 15 mins late; numbers of urgent journeys.</p>	Annual statistical return KA34 completed by ambulance service provider	Ambul. Trust	<input type="checkbox"/>	Current rural/urban classification used for ambulance services is classified urban if > 2.5 persons per hectare, otherwise rural. Breakdown into rural/urban areas may not be very sensitive at Ambulance Trust level as this covers large areas.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>NHS Performance Indicators: Annually published indicators for performance, which are measured for all NHS Trusts and PCTs in England and are converted into star ratings (ranging from zero to three stars). In 2002 CHI took over the responsibility from DH for assessing NHS organisation performance. The latest performance indicators available are for 2002/03. A large range of indicators form the core key targets which cover priority areas set by the Government, and also a wider range of other important areas covering capacity, clinical focus, patient focus, access to quality services, improving health indicators and service provision. The CHI National Patients Survey (see below) contributes to the performance indicators.</p>	<p>Data comes from various sources including ONS, a variety of DH statistical returns and Hospital Episode Statistics.</p>	<p>PCT/NHS Trust</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban breakdown between Trusts is used. Breakdown into rural/urban areas may not be very sensitive at Trust level as this covers large areas.</p>
<p>CHI National Patients Survey: There were 3 patient surveys carried out in 2003 by CHI: Emergency department patient survey (NHS Trust level – over 59,000 patients in 155 NHS Trusts in England); Outpatient survey (NHS Trust level – over 90,000 patients in 171 NHS Trusts) & National Patients survey level (PCT level – over 123,000 patients at 304 PCTs). Results contribute to the performance ratings for Trusts for 2002/03. Five areas were covered: 1. Access and waiting; 2. Better information, more choice; 3. Building relationships; 4. Clean, comfortable, friendly place to be; 5. Safe, high quality, co-ordinated care.</p>	<p>Annual patient surveys carried out at NHS Trusts and PCTs.</p>	<p>PCT/NHS Trust</p>	<p><input type="checkbox"/></p>	<p>No current breakdown of rural or urban areas. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas. However if it was changed to use postcodes then a definition of rurality could be used, but confidentiality is a priority which must be maintained in any breakdown.</p>
<p>Hospital Waiting Times: Data collected by DH for both PCTs and NHS Trusts – waiting times with numbers of patients/referrals by speciality and categorised waiting times for outpatients first attendances, ordinary and day case admissions, demand for elective admissions – events & deferred or suspended patients</p>	<p>Quarterly statistics of hospital waiting times from various DH statistical return forms</p>	<p>PCT/NHS Trust</p>	<p><input type="checkbox"/></p>	<p>No current breakdown of rural or urban areas. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Hospital Episode Statistics: Database held at DH, which includes detailed individual data on admitted patient care from 1989 to present. Data held includes patient details including identifiers and postcode; admission and discharge details; episodes and spells in hospital; diagnosis details; operative procedures carried out; organisations where treatment took place; practitioner/consultant details etc.	HES receives data from inpatient records from NHS hospitals in England (approx 12 million records)	Individual if research approved	<input checked="" type="checkbox"/>	Although rural indicators have been used in past, not used as standard within database. Not known if new classification would be used in future. Individual data would be available for bona fide research if approved by passing very strict requirements and confidentiality agreements signed. Data can be ordered from HES, and it may be possible that rural/urban indicators could be attached to data by HES team. Ward level data alternatively may be available. Some data down to NHS Trust/PCT level is available on DH website.
Hospital Activity Statistics: Data collected by DH and includes: bed availability – beds open overnight; day only beds; residential care beds; critical care beds; ward attendances; day care attendances; A+E attendances; outpatients attendances/Do Not Attends (DNAs); numbers of supporting facilities; imaging & diagnostics; cancelled operations; NHS written complaints.	Quarterly statistical collections from various DH statistical return forms	PCT/NHS Trust	<input type="checkbox"/>	No current breakdown of rural or urban areas. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Health Service Staff Characteristic/Recruitment				
NHS Staff census (medical and non-medical): Headcounts of all medical and non-medical NHS Staff, including dental staff, carried out by DH in annual censuses on 30 th September of each year.	Annual NHS workforce censuses carried out by DH.	PCT level	<input type="checkbox"/>	It is rare that such rural/urban splits in data needed but such breakdowns would be considered if required and found to be beneficial. However breakdown at Trust level may not be very sensitive as covers large area.
NHS Workforce Vacancy Survey: Annual collection by DH of 3 month vacancies (have lasted 3 months or more and which employers are actively trying to fill) as at 31 March – so focuses on recruitment difficulties. Data includes numbers of vacancies and vacancy rates for different staff groups and for each NHS Trust/PCT.	NHS Workforce Vacancy survey: Annual count of 3 month vacancies	NHS Trust/PCT	<input type="checkbox"/>	No current rural/urban breakdown. Breakdown into rural/urban areas may not be very sensitive at Trust level as this covers large areas.
GP Recruitment, Retention & Vacancy survey: Annual survey by DH of vacancies in general practice over annual period. Data includes: average number of applicants per vacancy; average time to recruit; difficulty in finding new GPs of sufficient quality; numbers of GPs leaving practices by sex and type of area.	Annual collection of GP vacancies in England and sample of 1,000 vacancies are analysed in detail.	PCT	<input checked="" type="checkbox"/>	Survey respondent self-defines the area that the practice serves out of: urban deprived; urban; mixed urban/rural; rural. However this is subjective and often not fully completed. Is interested in using new rural/urban definition if easily applied to practice postcode, as this will be independently defined from outset. Data only released to PCT level to maintain confidentiality.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Personal Social Services Staff & Expenditure: Data collected by DH. Staff data details numbers of social services staff by age group and ethnic origin. Expenditure data includes amounts spent for different category patients by type of care given.</p>	<p>Annual statistical returns: SSDS001 (staff numbers) and PSS EX1 (expenditure).</p>	<p>County/UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator or immediate plans to carry out a rural/urban breakdown. Aggregate data at LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas</p>
<p>Cancer</p>				
<p>Cancer Registrations: Data can be requested from ONS and registration numbers and rates to GOR level for all cancers are readily available in series MB1 publications. Data on numbers of new registrations, incidence, mortality and survival data for different types of cancer. Level of data available depends on how prevalent the cancer is. To ensure completeness of released data, there are delays until data are publicly available. Latest year for which data are publicly available is 2000.</p>	<p>Registrations of all malignant neoplasms – continuous collection</p>	<p>Individual if research approved.</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rurality indicator, but could include new classification by using postcode information if found to be advantageous. Ward level data should be available in near future for selected cancers – but content of this level data is still being discussed so it doesn't allow disclosure. Suppression carried out in publicly released data if count is less than 5 in any area. Individual level data are available if bona fide research approved and passes tight security requirements. Data for several aggregated years likely to be needed for meaningful research so that cancer incidence/prevalence numbers high enough.</p>
<p>Breast Screening Programme Statistical Bulletin: Data collected annually by DH. Includes the coverage (proportion of women resident with test in last 3 years) + uptake (proportion of women invited to screening for which screening test result is recorded) of breast cancer screening by age, test status and outcome.</p>	<p>Annual statistical return forms KC62 (screening unit level) & KC63 (PCT level)</p>	<p>PCT</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or plans to use new or other definition of rurality. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.</p>
<p>Cervical Screening Programme Statistical Bulletin: Data collected annually by DH and includes: coverage of screening by results of test; numbers of women invited in year by type of invitation; numbers of women tested in the year by type of invitation; numbers of different test results.</p>	<p>Annual statistical return forms KC53 (PCT level), KC61 (pathology lab level) & KC65 (clinic level)</p>	<p>PCT</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or plans to use new or other definition of rurality. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Mental Health				
Community Psychiatric Nursing: Data collected annually by DH and includes numbers of initial & first contacts by age group + sex by year.	Annual statistical return KC57	PCT	<input type="checkbox"/>	No current rurality indicator or plans to use new or other definition. Aggregate data to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Clinical Psychology Services: Data collected annually by DH and includes numbers of initial & first contacts by age group + sex, and by source of referral.	Annual statistical return KT24	PCT	<input type="checkbox"/>	No current rurality indicator or plans to use new or other definition. Aggregate data to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Guardianship under Mental Health Act 1983: Data collected annually by DH and includes numbers of cases opening and closing during year and average length of cases closed during year, and disorder type.	Annual statistical return SSDA 702	County/UA	<input type="checkbox"/>	No current rurality indicator or plans to use new or other definition. Anonymised case-level data collected from every LA. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.
Inpatients formerly detained under Mental Health Act 1983: Data collected annually by DH and includes numbers of admissions during year by MHA section number and mental category and sex; changes of legal status; Numbers of detained and informal patients by mental category and sex as at 31 March.	Annual statistical return KP90	PCT/NHS Trusts	<input type="checkbox"/>	No current rurality indicator or plans to use new or other definition. Aggregate data to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT/NHS Trust level as this covers large areas.
Community Learning Disability Nursing: Data collected annually by DH and includes numbers of first and initial contacts by age group and sex within year.	Annual statistical return KC58	PCT/NHS Trusts	<input type="checkbox"/>	No current rural indicator or plans to use new or other definition. Aggregate data to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Mental Health of Carers: Survey using people aged 16 or over and identified as carers in General Household Survey 2000. Data includes mental health of carers (measured by CIS-R scores) and characteristics of the carer; impact of caring on employment; characteristics of person cared for; characteristics of caring relationship; level of support received by carer; carer's receipt of treatment and health services; characteristics of people who are no longer carers; smoking, drinking and use of prescribed and non-prescribed drugs by carer.	Sample survey – sample obtained consists of 1,350 people in England.	GOR	<input checked="" type="checkbox"/>	Currently uses population per square hectare information and codes area as urban/rural/semi-rural based on interviewer's opinion. It is possible that other rural/urban indicators could be attached to data by ONS using postcode information, but this is a small sample, and breakdown would have to be checked that it is not disclosive.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Electro Convulsive Therapy 2002 survey: Data collected by DH as a 3 month snapshot and includes: numbers of persons per 100,000 population who underwent ECT by sex, age group, diagnosis; numbers of patients finishing ECT by sex and number of treatments administered. Similar survey was carried out in 1999 and comparison figures are given. Main objective of survey was to measure level of reporting of ECT in Hospital Episodes Statistics (HES) data – found to be under-reporting, but this is improving. Since 2003 a Mental Health Minimum Dataset has been set up – which will contain individual level derived from HES on the whole range of mental health treatments. As still early days this dataset is not yet robust.</p>	<p>Irregular survey of all NHS Trusts and some private facilities in England with mental health facilities on ECT administrations in 3 month period Jan-Mar 2002.</p>	<p>HRO</p>	<p><input type="checkbox"/> (ECT 2002 survey) <input checked="" type="checkbox"/> (New Mental Health Minimum Dataset)</p>	<p>No current rural/urban breakdowns used. Aggregate data at NHS trust level collected for this 2002 ECT survey. Due to relatively small numbers data are unlikely to be released publicly below HRO level. However if formally apply for data and research approved and respect stringent confidentiality agreements then survey data at NHS Trust level may be given. Breakdown into rural/urban areas may not be very sensitive at Trust level as this covers large areas. If research formerly approved as above, individual record data from the Mental Health Minimum Dataset (once robust) could potentially be used for analysis and could be attached to rural/urban indicators at small area level.</p>
Other Long-term Illness/Disability				
<p>People registered as deaf or hard of hearing: Data collected annually by DH and includes numbers of people on registers who are deaf or hard of hearing by age group as at 31 March. Since registration for deaf/hard of hearing is not compulsory, numbers will not give complete figures for England.</p>	<p>Annual statistical return SSDA 910</p>	<p>County/UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban breakdown. Aggregate data at LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.</p>
<p>People registered as blind or partially sighted: Data collected annually as at 31 March by DH and includes: numbers of people on registers who are blind or partially sighted by age group; numbers of new registrations in year by age group; numbers of registered blind people with additional disability by age group. Since registration for blind/partially sighted is not compulsory, numbers will not give complete figures for England.</p>	<p>Annual statistical return SSDA 902</p>	<p>County/UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban breakdown. Since registration is not compulsory, figures will not be complete. Aggregate data at LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.</p>
<p>Attendance allowance claimants: Data collected by DWP and gives all claimant counts at higher and lower rates in May. Latest data available is for 2001 using 2003 ward boundaries. 2002 and 2003 data will be available in 2004. Data for 1998, 1999 & 2000 is available at ward level (1998 boundaries) on Neighbourhood Statistics website.</p>	<p>Annual 100% scan of all claimants</p>	<p>Ward</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural/urban breakdown but may use new rurality definition if useful. Numbers of claimants at ward level is sufficient not to be disclosive.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Incapacity benefit & Severe disablement allowance claimants: Data collected by DWP. Incapacity benefit data gives claimant counts by age group, sex and age group of dependents in August. Severe disablement allowance data gives claimant counts by age group and sex in August. Latest data available is for 2001 using 2003 ward boundaries. Data for 2002 and 2003 will be available in 2004.</p>	Annual 100% scan of all claimants	Ward	✓	No current rural/urban breakdown but may use new rurality definition if useful. Numbers of claimants at ward level sufficient not to be disclosive.
<p>Disability Living Allowance: Data collected by DWP and gives all claimant counts by: care component rate; mobility component rate; sex and age group. Latest data available is for 2001 using 2003 ward boundaries. Data for 2002 and 2003 will be available in 2004. Data for 1998, 1999 & 2000 is available at ward level (1998 boundaries) on Neighbourhood Statistics website.</p>	Annual 100% scan of all claimants	Ward	✓	No current rural/urban breakdown but may use new rurality definition if useful. Numbers of claimants at ward level sufficient not to be disclosive.
<p>CJD Statistics: Data collected by the National CJD Surveillance Unit at DH. UK totals are released every month for the numbers of referrals of suspected CJD and the numbers of deaths occurring for all definite and probable cases of all types of CJD, for each year since 1990 (inclusive).</p>	Monthly data collected from a variety of clinical sources and also from death certificates obtained from ONS.	Individual if research approved	<input checked="" type="checkbox"/>	Data available in the public domain not usually broken down geographically to smaller than national levels. Numbers are very small, particularly for the definite CJD cases, and so disclosure is obviously the main concern. Small numbers will also affect the statistical significance of analysis results. Research using individual data would have to pass stringent confidentiality requirements. Various current research projects currently underway, including a case-control study investigating risk factors for the disease, and a study into geographically associated CJD cases.
Work-related Disease				
<p>HSE Epidemiology Medical Statistics Unit – Causes and Kinds of Disease: Various surveys and epidemiological studies analysed by industry including rural industries: agriculture, hunting, forestry and fishing. HSE admit that currently agriculture data are weakest but working to improve this. Includes data on pesticide use and related disease, and 'Farmer's Lung' (allergic alveolitis). Also Labour Force Survey models - Self-reported work-related illness survey data used.</p>	Various epidemiological studies. In past have used RIDDOR reporting forms but found these not reliable.	Industry type	<input checked="" type="checkbox"/>	No data currently publicly released but should be in near future. Data analysis is usually more concerned with nature of industry concerned rather than location. Anonymised individual data should be available in future. No current rurality indicators used but are interested in using new definition, and could be attached to postcode information when this is available.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Accidents				
HSE Injury statistics: Fatal and non-fatal injuries by industry, which include agriculture, hunting, forestry and fishing. Numbers and type of injury. Rates of injury based on ONS employment estimates. Labour Force Survey estimates also used to help determine level of national reporting of injuries. Numbers of non-fatal injuries are under-reported, particularly for self-employed.	Uses RIDDOR 95 form F2508. ONS employment and Labour Force survey injury estimates are also used.	County/ UA	<input checked="" type="checkbox"/>	No current rural/urban breakdown, but would use new rural/urban definition if found to be useful. However HSE work is really based on nature of industry rather than location.
HSE Enforcement Action Statistics: Enforcement notices and convictions - penalties /fines resulting. Data from convictions available on public register at individual business level.	All enforcement action convictions data updated weekly on website.	Business address	✓	No current rural/urban breakdown used, but new rural/urban definition could be readily added to data at business level.
Industrial Injuries Disablement Benefit: Data collected by DWP and includes: numbers of new claims for industrial injury prescribed disease or accident by prescribed disease/cause of accident; numbers of first diagnosed assessments for prescribed diseases by prescribed disease; numbers of claimants by age group; gender; percentage disablement & occupation; industrial injury claim decisions by type of claim and outcome.	Quarterly 100% scan of data from DWP local disablement benefit offices. Before Quarter 2, 2002 a 10% sample was used.	County/ UA level by request to DWP.	<input type="checkbox"/>	No current rural/urban breakdown, but would use new rural/urban definition if found to be advantageous. Numbers rounded to nearest 5 to protect confidentiality. Numbers relating to 10% sample (before Q2, 2002) will be subjected to sampling error. Data on industrial accidents can be released at LA level by request to DWP. Data on prescribed work-related disease only available to GOR level.
TARN (The Trauma Audit & Research Network): Database held at University of Manchester, which includes detailed data providing measurements on trauma care process (e.g. pre-hospital times, initial resuscitation, investigations details, times to theatre, interventions etc.) and outcome (e.g. death/survival/health-related quality of life post-trauma). Participating hospitals in TARN have access to all data. Data are rarely given out to 3 rd party researchers, but analysis can be requested and carried out by the TARN research team.	Data are collected from Hospital Information Systems. 50% of trauma-receiving hospitals participate.	ED	<input checked="" type="checkbox"/>	No current rural/urban breakdown. However information down to enumeration district is available to TARN researchers and so rural/urban classification could be attached at this level in the same way that deprivation indices are currently added to data. TARN can exclusively carry out research or allow researchers/clinicians to work within TARN. Would need to formerly apply with abstract and null hypothesis of proposed research. Writing up of research would need to be overseen by TARN research director.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Infant Feeding				
Infant Feeding Survey: Five-yearly survey, first taking place in 1975 and carried out by ONS from 1975-1995 and by BMRB Social Research in 2000 (latest survey). Data from 1985-2000 is available on UK Data Archive at the University of Essex. Data includes information on the birth & post-natal care; incidence, prevalence and duration of breastfeeding; mothers' intentions to breastfeed; age at introduction to solid foods; smoking & drinking during pregnancy.	2000 survey sample - 7382 births in England & Wales between 19/8/00 & 19/10/00 (9500 for whole UK).	Standard region	<input checked="" type="checkbox"/>	No rural/urban indicator currently held within dataset. It is possible that postcode information for 2000 survey could be de-archived and attached to rural/urban classification for analysis, at a charge. However, the possibility of incorporating rural/urban indicators into 2005 Infant Feeding survey design could be discussed. Coverage of sample over rural/urban areas would need to be investigated.
Sexual Health				
Contraceptive Services: Data collected annually and includes: numbers of female and male first contacts by age group and main method of contraceptive chosen; numbers of post-coital contraceptives given (females) by age group.	Annual statistical return KT31	PCT/NHS Trusts	<input type="checkbox"/>	No current rural/urban indicator or plans to use new or other definition. Aggregate data to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Abortion Statistics: Data collated annually by DH from 2002 (up to and including 2001 data was collated by ONS). Data includes numbers of abortions by purchaser; gestation period; percentage NHS funded under 10 weeks; age of woman. Case level data are collected in the notification form including marital status, parity, treatment details and if Chlamydia screening was carried out.	Annual collection of abortion notification form HSA4	Individual if research approved.	<input checked="" type="checkbox"/>	No current rural/urban indicator used or immediate plans to use new or other definition. Focus is more on where abortions take place (generally in urban areas) than where patients live. Individual data are available (with postcode information) if research approved by Chief Medical Officer and passes stringent confidentiality requirements.
National Survey of Sexual Attitudes & Lifestyles 2000: Survey carried out by NatCen, London School of Hygiene and Tropical Medicine and Royal Free and University College London Medical School. Similar survey also carried out in 1990. Data not currently available to public but will be put into UK Data Archive in next year or so. Data includes information on: general health and use of health services; sexual experiences; miscarriages; still births; terminations; drug use; HIV.	Sample survey of 11,161 individuals aged 16-44.	GOR (when data are publicly available)	<input checked="" type="checkbox"/>	No current rural/urban indicators included. When data deposited at UKDA, unlikely to be available below GOR level due to sample size and to maintain confidentiality. However new rural/urban indicator could be added by NatCen, for a charge, to individual postcode information, so that analysis between different rural/urban groups would be possible – as long as breakdown was not disclosive.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Smoking, Drinking and Drug use				
National Drug Treatment Monitoring System: Data held centrally by DH. Introduced in England in 2001. Formerly used separate regional drug misuse databases. Aggregate data at PCT level is held and includes: numbers of users starting agency episodes by gender and age group; injecting behaviour for various types of drugs in last 4 weeks of users starting agency episodes; agency episodes by type of agency. Also information collected for clients at end of each financial year on whether still in treatment. In future database will be web-based. Current discussions on possibly releasing some data to Neighbourhoods Statistics web site in future. In future database will be run by a special health authority called the National Treatment Agency for Substance Misuse.	Anonymised data regularly collected from drug treatment agencies by DH.	PCT	<input type="checkbox"/>	No current rural/urban breakdown used. No patient level information is held centrally in the database. Anonymised data available down to PCT level. Data at smaller levels may be available from data providers directly – e.g. local drug treatment agencies, GPs etc. Currently direct access to NDTMS data are only available to DH staff, NHS staff or staff from specific drug-related organisations. Not yet decided what access will be available for 3 rd party research.
Smoking, Drinking & Drug Use among Young People in England Survey: Annual survey carried out since 2000 by NatCen and NFER. Similar surveys carried out by ONS from 1982-1999. Data from 1990-2002 surveys are available from UK Data Archive at University of Essex. Sample should be representative of both rural and urban areas. Questions asked on respondents' experience of smoking, drinking and drugs; availability and purchase of cigarettes, alcohol and drugs.	2002 survey sample consisted of over 9,000 pupils in England aged 11-15, from over 300 schools.	School	<input checked="" type="checkbox"/>	No current rural/urban indicator used. NatCen could attach new urban/rural classification to school postcode data. However, no information on where individual pupils live, so sensitivity of this breakdown would need to be investigated.
Smoking Cessation Services: Data collected quarterly by DH. PCT level data includes numbers of people setting quit date and 4 week outcome. Data at SHA level includes breakdowns by age and sex; numbers receiving various treatments (e.g. Nicotine Replacement Therapy); rates of people who successfully quit per 100,000; changes in numbers of quitters since previous year.	Smoking Cessation Services Quarterly Monitoring Return sent to all SHAs.	PCT	<input type="checkbox"/>	No current rural/urban breakdown used and no immediate plans to use new or other rural/urban classification. Aggregate data mostly at SHA level, but some at PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level, and particularly at SHA level, as these cover large areas.
Drug Seizure & Offender Services: Annual statistical bulletin of data collected & analysed by the Home Office. Data includes numbers of seizures by type of drug; number of offenders given cautions; numbers of offenders taken to court and outcome. Clearly statistics only show drug offences caught by police and do not give true offence prevalence.	Seizure data from annual Crimsec38 form. Offender data from cautions and court appearance data.	Police Force Area	<input type="checkbox"/>	No current rural/urban breakdown. Aggregate data at Police Force Area level. Breakdown into rural/urban areas at this level may not be sensitive as cover large areas. Caution should be taken in rural/urban comparisons, as sparser police presence in the larger rural areas may result in lower offence detection rate than in urban areas.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Other Disease/Illness Related Statistics				
<p>Communicable Diseases: Database held by the Communicable Disease Surveillance Centre (CDSC) at the Health Protection Agency (HPA) of numbers of unconfirmed Notifications of Infectious Diseases (NOIDS), some of which are confirmed by laboratory tests. Main uses for the data are to help to detect outbreaks as they occur. Results are also collected from laboratory tests for communicable diseases.</p>	<p>Statutory notifications of unconfirmed infectious diseases that are sent to the CDSC by each LA. Results from laboratory tests are also sent to CDSC</p>	<p>CD/UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicators used. Weekly updates of NOIDS (unconfirmed) cases are publicly available at LA level every week, however as they are unconfirmed, these numbers are subject to both over- and under-reporting. To obtain other laboratory data, one would need to formerly apply for the data and have research approved. Although HPA do have resources to use postcode information, the postcode information from laboratory data are of poor quality, so data are not currently available at lower levels. However HPA are pushing to try and improve laboratory address data quality.</p>
<p>Environmental Statistics: Number of different types of datasets of environmental interest collated and summarised by Defra. Datasets which might possibly be of interest in looking at geographic variations such as rural/urban differences are concerned with air quality; bathing waters; drinking water; radioactivity; waste development</p>	<p>Number of different datasets concerned with monitoring environmental factors</p>	<p>Monitoring sites/regions</p>	<p><input checked="" type="checkbox"/></p>	<p>The majority of the data collected are at a regional or national level, or are collected from very specific areas for which rural/urban splits may not be sensible such as coastal/marine waters. Data collected on air quality are collected at a number of monitoring sites across the country. The monitoring sites are classified as kerbside, roadside, urban centre, urban background, suburban, rural, remote, and special emission source, and so rural/urban differences could perhaps be measured here. Could possibly attach new rural/urban classification to monitoring sites.</p>
<p>UK Transplant Data: National Transplant Database held by UK Transplant (a special Health Authority within the NHS which was formed in 1991) which holds comprehensive individual data on organ recipients' waiting time, clinical details about transplant operation, and follow up after transplant. Individual data includes recipient's postcode information. Data are also held on which hospital the donor is from, but individual postcode information is not held for the donor. UK Transplant also holds the Organ Donor Register database holding details on all people who have signed up to be an organ donor.</p>	<p>Database including details of all transplants in the UK within the NHS</p>	<p>Individual data if research approved</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural/urban breakdown. However postcode information is held for the recipients' residence, and so it is possible that the new urban/rural classification could be attached at postcode level. Some 3rd party research is done using the database. Annual data at NHS Trust and Transplant Centre level is released on the website. To obtain data for research at a smaller geographical level would have to formally apply and the proposed research protocol would have to be approved.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Congenital Anomaly Statistics: Annual collation by ONS of congenital anomaly notification statistics – produced in annual MB3 publication. Currently, 2001 data are the latest published. Includes all congenital anomaly notifications for both live and stillborn babies except those affected by a given list of minor anomalies which ONS do not need to be notified of and are excluded</p>	<p>Continuous collection of SD56 notification form.</p>	<p>SHA</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural/urban breakdown. After NHS reorganisation, data will now be published to Strategic Health Authority level – data was published to smaller former health authority areas up to 2001. Possible breakdowns in future using new rural/urban classification would depend on numbers involved. Data are available for researchers if research is approved and confidentiality ensured. However, anonymised individual data can only be released with permission of data custodians who originally provided data.</p>
<p>Other Health Services Statistics</p>				
<p>Chiropody services: Data collected annually by DH and includes: numbers of initial contacts by age group; total number of first contacts throughout financial year.</p>	<p>Annual statistical return KT23</p>	<p>PCT/NHS Trust</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or immediate plans to use the new or other rurality definition. Aggregate data to Trust level. Breakdown into rural/urban areas may not be very sensitive at Trust level as this covers large areas.</p>
<p>Community Care - Adult Home Care: Data collected by DH and includes: contact hours provided & households receiving home help/care by sector of provider; intensive home care; total contact hours for each sector of provider; number of visits for each sector of provider; contact hours of home help & home care per 10,000 households by sector of provider; households receiving home help & home care per 10,000 households by sector of provider.</p>	<p>Annual statistical return HH1.</p>	<p>County/ UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban breakdown used or immediate plans to use the new rural/urban classification. Aggregate data to LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.</p>
<p>Community Care - Private Nursing Homes, Hospitals & Clinics: Data collected annually by DH and includes at PCT level: registered nursing beds & intended use as at year end; homes, beds & staffing 2000-2001. SHA and HRO level data includes: numbers of registered premises by premise type since 1996/97; numbers of registered nursing beds by premise type since 1996/97; qualified nursing types by premise type; numbers of occupied nursing beds by age group & client group; numbers of nursing staff in post by grade.</p>	<p>RH(N) A, RH(N) B and RH(N) S annual statistical return forms.</p>	<p>PCT</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban breakdown used, or immediate plans to use new rural/urban classification. Aggregate data to SHA and PCT levels. Breakdown into rural/urban areas may not be very sensitive at Trust level as this covers large areas.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Community Care - Local Authority Supported Residents (Adults): Data collected by DH and includes: numbers of LA supported residents at 31st March: - by type of accommodation; in LA and registered care homes by client group and type of care; in permanent residential & nursing care by age group & type of care; in temporary residential & nursing care by age group and type of care; LA supported admissions in permanent residential & nursing care by client group and type of care; LA supported admissions in permanent residential & nursing care by client group & type of care.</p>	Annual SR1 statistical return forms completed by all LAs	County/ UA	<input type="checkbox"/>	No current rural/urban indicator and no immediate plans to use new or other rural/urban definition. Aggregate data to LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.
<p>Prescription Cost Analysis: Annual data collection by DH. Data includes: all prescription items dispensed in community within England; numbers of prescriptions charged for, and by categories for which no prescription charge is made; numbers of prescriptions by category of exemption of NHS Low Income Scheme; Net Ingredient Cost of prescriptions made.</p>	Data are taken from Prescription Pricing Authority (PPA) information systems	PCT	<input type="checkbox"/>	No current rural/urban indicator or immediate plans to use new or other rural/urban definition. Data are available to PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
<p>Community Pharmacies in England & Wales: Annual collection of data by DH on community pharmaceutical services in England & Wales. Data includes: numbers of pharmacies in contract; openings and closures of pharmacies in year; payments received for additional services in year; collection & disposal of unwanted medicines;</p>	Annual statistical return PHS1.	PCT	<input type="checkbox"/>	Areas are defined as 'controlled' if deemed to be predominantly rural. Controlled areas hold more restrictions on which community pharmacies can open – due to less room for competition. Population density (per sq km) for each PCT is also calculated. No immediate plans to use new or other rural/urban definitions. Data are only available at PCT level.
<p>District Nursing: Data collected by DH and includes numbers of first contacts with district nurses by age group and sex, and numbers of initial contacts during financial year.</p>	Annual statistical return form KC56	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new rurality definition. Aggregate data to Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
<p>Health Visiting: Numbers of first contacts with health visitors by age group during financial year.</p>	Annual statistical return form KC55	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new rurality definition. Aggregate data to Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
NHS Immunisation Statistics: Data collected by DH and includes: numbers of boosters/reinforcing doses given in year to school leavers aged 13-16 only for Diphtheria, Tetanus & Polio; numbers of courses of MMR completed in year for children aged 13-16; numbers of BCG skin tests found positive/negative in year by age group; numbers of BCG vaccinations in year by age group.	Annual statistical return form KC50	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new or other rurality definition. Aggregate data at Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Occupational Therapy Services: Data collected by DH and includes: numbers of initial contacts during year by source of referral; numbers of initial contacts during year by age group & sex.	Annual statistical return form KT26	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new or other rurality definition. Aggregate data at Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Physiotherapy services: Data collected by DH and includes: numbers of initial contacts during year by source of referral; numbers of initial contacts during year by age group & sex	Annual statistical return form KT27	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new or other rurality definition. Aggregate data at Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Specialist Nursing Care: Data collected by DH and includes: numbers of initial contacts during year for each of: Macmillan nurses; hospice nurses; Marie Curie nurses; stomatherapists; continence Nurses; premature baby nurses; diabetes nurses; oncology nurses; other terminal/cancer care nurses; catheter nurses; breast care nurses; community paediatric nurses.	Annual statistical return form KC59	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new or other rurality definition. Aggregate data at Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
Speech & Language Therapy: Data collected by DH and includes: numbers of initial contacts during year by source of referral; numbers of initial contacts during year by age group & sex.	Annual statistical return form KT29	PCT/NHS Trust	<input type="checkbox"/>	No current rural/urban indicator used and no immediate plans to use new or other rurality definition. Aggregate data at Trust level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas.
General Ophthalmic Services: Data collected annually by DH and includes numbers of sight tests and vouchers reimbursed per 10,000 population; Numbers of different Optometrists and Ophthalmic medical practitioners per 10,000 population.	Data are collected from the NHS workforce censuses and a sample of optical vouchers	PCT	x	No current rural/urban breakdown. Data are only available at PCT level. Breakdown into rural/urban areas may not be very sensitive at PCT level as this covers large areas. However, more than half of ophthalmic practitioners hold a contract with more than one PCT and so comparing ophthalmic services workforce between PCTs may not be so useful. Data from voucher survey not available for further research.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<i>Other National Surveys with Health-related Data</i>				
<p>Health Survey for England: Cross-sectional survey carried out jointly by NatCen and the Department of Epidemiology and Public Health at University College London, on behalf of DH. Data currently available up to 2001 survey on UK Data Archive at University of Essex. Questions covered in survey include general health and long-standing illnesses; use of health services; drinking & smoking; psycho-social health; non-fatal accidents. Survey focuses on different particular health issues in different years e.g. cardiovascular disease & risk factors covered in 1991-1994 and atopic disease (e.g. asthma, eczema etc.) covered in 1995-96.</p>	<p>1991-ongoing cross-sectional survey. 2001 survey sample consisted of 19640 individuals in England</p>	<p>Former HA level for 2001 survey. PCT level may be available from 2002 survey, when this data are released.</p>	<p><input checked="" type="checkbox"/></p>	<p>An 'urban' variable is currently included within the dataset and is coded out of urban/suburban/rural. However this is derived from question answered by interviewers' opinion of residence location, and is therefore subjective. Ward level data may be available if 2-3 years worth of data was aggregated – would need to request in writing for such small area data, and research would need to be approved by DH. Would be possible, for a charge, for NatCen to release data with new rural/urban classification if they could easily attach indicator to the postcode information they hold.</p>
<p>ELSA (English Longitudinal Study of Ageing): Longitudinal study being carried out by NatCen, International Centre for Health and Society at University College London and the Institute of Fiscal Studies. Data from first wave should be available on UK Data Archive at University of Essex in first half of 2004. Each member of the study will be interviewed every 2 years. Questions asked designed to explore in particular the relationships between health and wealth, social network and support, family and household composition, nature of retirement. Plan is to link data to the Health Survey for England. A parallel study is also being carried out in the US and comparisons will be made between the 2 countries.</p>	<p>Longitudinal study with sample consisting of approx 12,100 adults (and about 100 partners) in England from Health Survey for England sample aged 50 and over.</p>	<p>Current discussion on what geog. unit will be available</p>	<p><input checked="" type="checkbox"/></p>	<p>Currently a rural/urban indicator may be used, but if so, person interviewed was not sure what the classification used was. Survey research team were aware of new rural/urban classification and current discussions will include whether to use this new rural/urban classification in the future within the study design for future waves. Discussions currently taking place on what the smallest geographical unit available will be.</p>
<p>Omnibus Survey: Monthly stratified random sample survey carried out by ONS – includes core questions that are asked every month, alongside non-core questions that change every month. Data currently available up to July 2002 on UK Data Archive at the University of Essex. Regular health modules carried out look at contraception, smoking, drinking. Also several months worth of surveys have included modules on anti-infectives and cancer screening/awareness.</p>	<p>Survey made up of a number of different modules that are available separately. Approx 1800 individuals aged 16 or over sampled every month</p>	<p>GOR</p>	<p><input checked="" type="checkbox"/></p>	<p>Currently population density per square hectare is used to form rural/urban indicator if needed. However rural/urban classifications not routinely attached to all datasets. ONS survey research team were aware of new classification, but no immediate plans to use this. However, likely that this will be discussed in future when classification is finalised.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>ONS Longitudinal Study: Longitudinal study carried out by ONS since 1971 census. Data are held by CeLSIUS (Centre for Longitudinal Study Information and User Support) and data are available for academic research from them. The census data available includes: age; sex; marital status; household composition; accommodation & tenure; country of birth and parent's country of birth; ethnicity (1991 & 2001); education and qualifications; economic activity; occupation and social class; migration and travel to work; long-standing illness (1991 & 2001) and self-reported health (2001); religion (2001); care-giving (2001); marriage and fertility history (1971). Events data linked to census data includes live and still births; infant deaths; cancer registrations; mortality and cause of death; widower status etc. from information on registered births/deaths/cancers sent to ONS.</p>	<p>Longitudinal study using census data from 1971 census onwards with intra-censal events linked for about 1% population in England and Wales. Census data are also available for all other co-residents of LS study sample members.</p>	<p>Grouped ward level if not disclosive.</p>	<p><input checked="" type="checkbox"/></p>	<p>Rural/urban indicator is included in dataset but is derived from 1971 census and so is now out of date. A problem of using rural/urban indicators in longitudinal study is that rural/urban areas change over time, particularly since 1971. The new rurality definition could be added at postcode level and this would be done by ONS staff within ONS safe setting. Ward level data could not be released, but data at higher levels if not disclosive could be released. CeLSIUS website includes useful training modules on how to use the ONS Longitudinal Study data.</p>
<p>British Household Panel Survey: Longitudinal study carried out by the Institute of Fiscal Studies and the Institute of Social and Economic Research. Data available from 1991-2001 (waves 1-11) in the UK Data Archive at the University of Essex. Data focuses on individuals' labour force experiences since leaving full-time education. Data includes information on household composition; employment & earnings; social support; accommodation & tenure; neighbourhood; individual demographics; residential mobility; health & caring; general health; lifetime childbirth; attitudinal items; activities of daily living. Past specific projects have linked BHPS data to NOMIS labour force data using Travel to Work areas</p>	<p>The total UK sample covers approximately 10,000 (England: 4,000) households and 20,000 (England: 8,000) individuals.</p>	<p>UA/CD</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rurality indicator used, and research team were not aware of the new rural/urban definition. Would use new classification in future if found to be advantageous. Sample is selected to be nationally representative and so sample sizes in rural areas are likely to be relatively small compared to sample in urban areas. It is likely that sample size is too small to compare intra-rural variations. Linkage of new rurality definition to postcode information could be carried out by researchers at the Institute of Social & Economic Research, and then data could be released, if resulting breakdown was not disclosive.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>General Household Survey: Annual cross-sectional study carried out by ONS of households in Great Britain. It has been carried out annually since 1971 except for breaks in 1997/98 and 1999/2000. Includes a household questionnaire and an individual questionnaire answered by all adults 16+ within sampled households. Data are available up to 2001/02 survey at the UK Data Archive at the University of Essex. Data includes information on household composition; tenure and accommodation; consumer durables; employment; income & pensions; education; health & use of health services; smoking & drinking; family information including marriage, cohabitation and fertility; income; demographic information including migration.</p>	<p>2001 survey consisted of an obtained sample of 8989 households and 21180 people (Great Britain)</p>	<p>GOR</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural/urban breakdown used. The survey research team were aware of the new rural/urban definition but have no immediate plans to use this or other classifications. However, the possibility of using new classification is likely to be discussed in the future when it is finalised.</p>
<p>Family Resources Survey: Annual cross-sectional study using multi-stage stratified sample carried out by DWP, ONS and NatCen. Principle aim of study is to provide monitoring of programme for social security and benefits. Data are currently available from 1993-94 to 2001-02 at the UK Data Archive at the University of Essex. Data are available on: occupation & employment; state benefits; NI contributions; interests & dividends; investments. Other information given includes household characteristics; accommodation and household costs; children's health; health restrictions on work; use of health services.</p>	<p>Sample covers around 24,000 households per year – 2001/02 sample obtained was of 25,320 households (Great Britain)</p>	<p>GOR</p>	<p><input checked="" type="checkbox"/></p>	<p>No rural/urban breakdowns currently used – additions are likely to be discussed in future. However, a local authority indicator can be released to use in data analysis for approved research, but not sure how robust this is. Data release below GOR level would need to be approved by DWP, NatCen and the Microdata Release Panel within ONS. Postcode information could be used to attach new rural/urban classification to existing datasets, within either NatCen or ONS. Level of rural/urban breakdown released would need to be sufficient to prevent disclosure.</p>
<p>British Social Attitudes Survey: Annual cross-sectional study carried out by NatCen . Data from 1983-2001 currently available at the UK Data Archive at the University of Essex. Social issues covered include defence; economy; labour market participation; welfare state; political attitudes; public spending & welfare; health care; understanding of public policy; health & safety of workplace; attitudes and use of health services; national identity; education; drugs and use; transport; demographic data. Questions about health service are only asked to a random two-thirds sample (2002 survey: 1924 individuals).</p>	<p>2002 sample consisted of 2897 individuals (18+) in England.</p>	<p>Ward level</p>	<p><input checked="" type="checkbox"/></p>	<p>The survey datasets only current measure of rurality is by population density. The new rural/urban classification could be attached by NatCen, for a charge, to the dataset using postcode information and then released. Sample sizes used for health-related questions may be too small for useful rural/urban analysis.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Labour Force Survey: Quarterly survey carried out by Labour Market Statistics Group at ONS. Survey was bi-annual from 1979-83, annual from 1984-1991 and quarterly from 1992 to present. Five waves interviewed every quarter, and each wave is interviewed for 5 successive quarters, so that as well as cross-sectional data, 5-quarter longitudinal data are also available. Data are available in quarterly datasets and also in annual datasets. Data includes information on individual demographics; family characteristics; full time education; employment and occupation; employees; Self employment; unemployment; ethnic minorities; qualifications; job-related training; health and effects on work. Not all variables are available to LA district level. All data available at the UK Data Archive at the University of Essex. Quarterly and annual datasets are available to download from the NOMIS database. Self-reported Work-related Illness surveys are derived from Labour Force Survey data and used by HSE.</p>	<p>Approx. 55,000 households respond every quarter equals approx. 150,000 individuals (about 0.2% population in Great Britain).</p>	<p>UA/CD – however not all data are available at this level.</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural indicator used. Survey research team were aware of the new definition. In theory the new classification could be added to the dataset using postcode information but would only be included if thought to be beneficial and found to work. When new definition available, the possibility of adding this indicator to the dataset is likely to be discussed. Ad hoc additions to existing datasets for specific projects are not routinely done, due to the strain on resources.</p>
<p>Expenditure & Food Survey: Annual survey carried out by ONS and Defra. 2001/02 survey combines Family Expenditure Survey (FES) and National Food Survey (NFS) for the first time. Data included in the UK Data Archive for 2001/02 currently only includes Expenditure component. Food component should be deposited by Defra when available. Both separate FES and NFS surveys are available at the UK Data Archive for earlier years. Data includes a household questionnaire about detailed household purchases; one-time & regular payments made. Each 16+ individual in households is also asked to keep expenditure diary for 2 weeks. Health data includes health insurance spending and expenditure on health services and prescriptions, medical products, spectacles etc.</p>	<p>2001/02 sample included 7473 responding households (UK).</p>	<p>GOR</p>	<p><input checked="" type="checkbox"/></p>	<p>Current rural/urban indicator used is the ODPM definition on 1991 urban settlement areas (based on the continuous built-up areas). It is likely that new rural/urban definition will be used once finalised.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>UK 2000 Time Use Survey: First large-scale study of its kind in the UK. Carried by ONS and Ipsos-RSL. Data are held in the UK Data Archive at the University of Essex. A previous Time Use survey funded by ESRC was carried out in 1987. At each responding household a household questionnaire was completed. Each individual aged 8 and over was asked to complete an individual questionnaire, 2 one-day diaries and 1 week work and education time sheet. Data consists of very detailed and comprehensive time-use activities. Includes data on general health and health restrictions on activities.</p>	<p>Obtained survey sample included 6414 households. Next full Time Use survey scheduled for 2010.</p>	<p>GOR</p>	<p><input checked="" type="checkbox"/></p>	<p>Population density fields are used, as well as fields with mosaic data (describes lifestyle type of area) and fields for proximity to nearest facilities. Survey research team would be very interested in adding the new rural/urban classification when it is available and this could be attached using postcodes or grid referencing so that different rural/urban groups could be compared. The smallest geographical unit that is currently available is GOR level to maintain confidentiality.</p>
WIDER DETERMINANTS OF HEALTH				
Child Care				
<p>Children Adopted from Care: Data collected annually by DH and DfES. It includes numbers of children adopted during year who were previously cared for by an LA council in England. Overall England breakdowns also include data on: durations of time until adoption and between stages in adoption process; age at adoption; gender of child and whether it is adoption by foster carer.</p>	<p>Annual statistical return AD1.</p>	<p>County/ UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or immediate plans to use new or other classification. Anonymised individual data collected from each LA. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas. For confidentiality reasons data are rounded to nearest 100, if exceeds 1000, and to nearest 10 otherwise. Numbers from 1 to 5 are suppressed.</p>
<p>Child Protection Register: Data collected annually by DH and DfES and includes: numbers and rates of referrals, initial and core assessments during year; numbers of registrations, re-registrations & de-registrations during year; total numbers of children on register as at 31st March by age group and sex; numbers of children on registers by category of abuse;</p>	<p>Annual statistical return CPR3.</p>	<p>County/ UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or immediate plans to use new or other classification. Aggregate data to LA level. Breakdown into rural/urban areas may not be very sensitive at LA district level as this covers large areas. Suppression of numbers less than 5 and rounding carried out to ensure no disclosure.</p>
<p>Children looked after by Local Authorities: Data collected annually by DH and DfES and includes detailed information on numbers and rates of children looked after in LA organisations by: age and sex; legal status; category of need; duration of being in care. Also some data on numbers of children no longer in care and activity of children on 19th birthday of care leavers.</p>	<p>Annual statistical returns CLA 100, SSSA 903, OC1, AD1 and OC3.</p>	<p>County/ UA</p>	<p><input type="checkbox"/></p>	<p>No current rural/urban indicator used or immediate plans to use new or other classification. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas. Suppression of numbers less than 5 and rounding used to ensure no disclosure. Samples are used for returns SSSA 903 and OC3 and so some data from these forms will include sampling errors.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Children's Homes: Data collected triennially by DH – last data available for 2000. Data as at 31 st March from CH1 return includes type of homes, facilities provided, age criteria for accommodation and maximum capacity of children that can be accommodated. Data on numbers of children placed in homes at 31 st March estimated from SSDA 903 return.	Triennial collection from statistical returns CH1 and SSDA 903	County/ UA	<input type="checkbox"/>	No current rural/urban indicator used or immediate plans to use new or other classification. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas. Small, unregisterable homes are excluded (less than 4 children, and private organisations run for profit) – in future these homes will also be registered and included in these statistics.
Children Accommodated in Secure Units: Data collected annually by DH and DfES and includes numbers of approved places and numbers of children accommodated at 31 st March by: sex; age group; length of stay; current legal status.	Annual statistical return SA1.	Secure Unit	×	No current rural/urban breakdown used. However, only aggregate data at secure unit level collected and only 31 secure units in England and Wales. No data on child's former place of residence, so not very relevant data set to analyse for rural/urban differences.
Children's Day Care Facilities: Data collected annually by DfES and includes information on following data provided by each LA as at 31 st March: day nurseries; playgroups and pre-schools; out of school clubs and holiday schemes for school-age children; family centres.	Statistical returns completed by all LAs	County/ UA	<input type="checkbox"/>	No current rural/urban indicator used or immediate plans to use new or other classification. Aggregate data at LA level. Breakdown into rural/urban areas may not be very sensitive at LA district level as this covers large areas.
Child Support Agency Statistics: Quarterly statistics collected by DWP. Data includes numbers of qualifying children by average maintenance payment; area of residence and employment/benefit status of both Person With Care and Non-resident Parent. In future, a more accurate 100% scan of the data should be available.	Quarterly 5% scan of Child Support Computer System	County/ UA level by request to DWP.	<input checked="" type="checkbox"/>	No current rural/urban indicator used or immediate plans to use new or other classification. Postcode information is held so attachment to data of rural/urban indicator is possible if found to be useful and numbers involved are large enough. Considerable sampling errors involved as using 5% sample, so figures under 500 should only be taken as broad estimate.
Child Benefit Statistics: Annual scan of Child Benefit Computer System by DWP up to 2003, and by Inland Revenue from May 2003. Latest data available is for 2001 at 2003 ward boundaries on DWP website. Ward level data includes total numbers of children claimed for, by age group of child, gender of child, and number of families by number of children in family. Ward level data for 1999 and 2000 is available on ONS website in Neighbourhood Statistics using 1998 ward boundaries.	Annual 100% scan of all claimants.	Ward	✓	No current rural/urban indicator. New rural/urban classification could be easily applied to ward level data. However, virtually all families with children in the UK receive child benefit, which is not affected by family income or savings. Therefore it is not a good indicator of economic differences, but can be used as an indicator for families with children.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Maternity Allowance: Quarterly statistics collected by DWP. Data includes: numbers of maternity allowance spells by weekly amounts paid throughout quarter; numbers of current maternity allowance spells as at end of quarter.</p>	<p>Quarterly 5% scan of all claimants.</p>	<p>County/ UA</p>	<p><input checked="" type="checkbox"/></p>	<p>No current rural/urban indicator. Statistics only broken down as standard to GOR level. LA level data are available, but release of data would depend on numbers involved, and if too small would be suppressed. Only 5% sample used so there will be considerable sampling errors, and therefore numbers should only be used as a rough guide. Postcode information is held so theoretically possible that rural/urban indicator could be attached to data, but release of breakdown would depend on numbers involved, and accuracy of postcode information available.</p>
Crime				
<p>British Crime Survey: Annual survey (biennial until 2000) carried out by the Home Office and NatCen up to 2000. BMRB International took over from NatCen from 2001 survey onwards. Sample size increased considerably from 2001 survey. Currently, data up to 2001 survey is included in the UK Data Archive at the University of Essex. Data includes detailed information on each respondent's experience of crimes, perceptions of their neighbourhoods; contact with police; attitudes to police; drug knowledge and use; sexual victimisation and rape etc.</p>	<p>2001 sample consists of approximately 40,000 interviews from England and Wales throughout year, including an ethnic booster.</p>	<p>Police Force Area</p>	<p><input checked="" type="checkbox"/></p>	<p>ACORN topology codes are included in the dataset and have been used by Home Office for rural/urban comparative analysis. Survey research team were aware of the new classification, and would be interested in using new definition if found to be advantageous. If research approved, definition could be added on an ad hoc basis to postcode level data from earlier surveys.</p>
<p>Youth Lifestyles Survey: 2 surveys carried out by the Home Office with Market and Opinion Research International (MORI) in 1992/93 and NatCen in 1998/99. Both survey datasets are held at the UK Data Archive at the University of Essex. Data covers information on schooling & qualifications; work/training/unemployment; income & expenditure; family life; housing & household structure; leisure activities; worries about crime, world and personal events; community safety and attitudes; victimisation experiences; attitudes to crime and punishment; contact with police.</p>	<p>1998/99 sample consisted of 4,848 young people aged between 12-30 living in England & Wales based on 1998 British Crime survey sample.</p>	<p>Standard Region</p>	<p><input checked="" type="checkbox"/></p>	<p>Uses ACORN codes within dataset, which can be used to distinguish between rural and urban area types. NatCen could add new rural/urban definition to the postcode information it holds on the YLS data on an ad hoc basis, for a charge. This particular survey was conducted quite a few years ago. However, likely to discuss adding new definition to new, upcoming Home Office Crime and Justice Survey (CJS) which will have a young person (aged 10-25) component and will be more up to date. Interviews for CJS will take place in 2004, 2005, and 2006.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Probation Statistics: Data collected by the Home Office and includes: numbers of persons starting community rehabilitation orders; numbers of persons starting community punishment orders; numbers in ethnic minority groups starting community punishment orders; numbers of drug treatment & testing orders; percentage of probation staff by ethnic minority group. Currently latest data available is for 2001. 2002 data will be available from 21st Jan 2004.</p>	Data collected from statistical returns made by the Probation Service on court proceedings.	Police Force Area	<input type="checkbox"/>	No current rural/urban breakdown used and no immediate plans to use new or other rural/urban definitions. Data from Home Office is only available down to Police Force Area, which covers quite a large area, and so a rural/urban breakdown at this level will not be very sensitive.
<p>Motoring Offences & Breath Tests: Data collected by the Home Office and includes: numbers and percentages of offences dealt with by written warnings, vehicle defect rectification scheme, fixed penalty notices and court proceedings by offence group; numbers of screening breath tests with number positive or refused; proceedings at magistrates courts by offence group; findings of guilt at all courts by offence group. Clearly statistics only show motoring offences caught by police and do not give true offence prevalence.</p>	Data on motoring offences and breath tests provided by police forces. Data on penalty charge notices provided by local authorities.	Police Force Area	<input type="checkbox"/>	No current rural/urban breakdown used and no immediate plans to use new or other rural/urban definitions. Data from Home Office is only available down to Police Force Area, which covers quite a large area, and so a rural/urban breakdown at this level will not be sensitive. Caution should be taken in rural/urban comparisons, as sparser police presence in the larger rural areas may result in lower offence detection rate than in urban areas.
Labour Market/Unemployment				
<p>NOMIS Official Labour Market Statistics: Web-based database held at Durham University on behalf of ONS. Data held consists of: annual business enquiry (and earlier equivalents) employee analysis 1981-2001 down to wards; 1991 census workplace statistics down to wards; claimant counts (unemployment-related benefits) from Jun 1983 to Dec 2003 down to wards; employee job estimates and job densities from 1981 to 2003 down to local authorities; Jobcentre Plus vacancies 1980 to 2003 down to jobcentre areas; Labour Force Survey annual and quarterly estimates from 1992 to 2003 down to local authorities; New Earnings Survey estimates from 1999 to 2003 down to wards; midyear population estimates from 1981 to 2002 down to local authorities; VAT registrations & Stocks from 1980 to 2002 down to local authorities. Much of the NOMIS data are likely be integrated into the Neighbourhood Statistics website in the future.</p>	Data collected from range of surveys and other statistics concerned with labour market characteristics	Ward	✓	No current rural/urban breakdowns in data held on the database. However, the new rural/urban classification should be readily attached to the ward level data.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>June Agricultural and Horticultural Census: Annual June collection of data from agricultural holdings carried out by Defra. Census data for 1990, 1995, 2000, 2001 & 2002 downloadable from Defra website down to ward level. Became a sample survey covering approximately 60% of all farms in 1995, rather than a census. However a census is still carried decennially (last time in 2000). June data includes areas of land owned & rented; changes in land area; areas of land used for specific types of crops; numbers of different types of animal stock; details of people working on holding & types of contract work given. A December survey provides further information on livestock numbers changes from June census and details of winter plantings etc.</p>	<p>Census (100% of holdings surveyed) until 1994. From 1995 has been sample survey (approx 60% holdings sampled), except for every 10th year when 100% census completed (last time in 2000).</p>	<p>Ward</p>	<p>✓</p>	<p>Currently uses a rural/urban indicator created by the Countryside Agency and Oxford University which colour codes ward level map data. Defra are very enthusiastic about using the new classification when finalised. This could be readily applied to ward level data. Grid referencing of agricultural holdings is currently used to allocate holdings to wards. Some suppression of data is required at ward level to protect confidentiality.</p>
<p>Jobseekers Allowance: Annual scan of all claimants by DWP. Latest data available is for 2001 using 2003 ward boundaries on DWP website. Ward level data includes total number of claimants by benefit entitlement, by age group and by age of dependents. Ward level data for 1998, 1999 and 2000 is available on ONS website in Neighbourhood Statistics using 1998 ward boundaries.</p>	<p>Annual 100% scan of all claimants</p>	<p>Ward</p>	<p>✓</p>	<p>No current rural/urban breakdown but may use new rurality definition if useful. Numbers of claimants at ward level is sufficient not to be disclosive.</p>
<p>New Deal for Young People (18-24) & Long-term Unemployed 25+: New deal is a scheme to provide people on unemployment benefits the help, support and training to get into work. Statistics published quarterly by DWP include information on numbers of participants in New Deal, and immediate destinations on leaving New Deal scheme. DWP are currently discussing the release of ward level data for Neighbourhood Statistics, part of ONS website.</p>	<p>Quarterly release of 100% data on participants in New Deal schemes</p>	<p>County/ UA level by request to DWP</p>	<p>☑</p>	<p>No current rural/urban indicator used but interested in using new classification and likely to discuss this with ONS in future. Ward level data potentially available in future, but data quality of postcode information held needs to be improved and DWP currently working on this. Content of ward level data depends on numbers involved to ensure no disclosure.</p>
<p>New Deal for Lone Parents: This scheme is to provide help for lone parents who want to work to find a job suiting their lifestyle. Statistics published quarterly by DWP include information on numbers of participants in New Deal, and immediate destinations on leaving New Deal scheme. DWP are currently discussing the release of ward level data for Neighbourhood Statistics, part of ONS website.</p>	<p>Quarterly release of 100% data on participants in New Deal schemes</p>	<p>County/ UA level by request to DWP</p>	<p>☑</p>	<p>No current rural/urban indicator used but interested in using new classification and likely to discuss this with ONS in future. Ward level data potentially available in future, but data quality of postcode information held needs to be improved and DWP currently working on this. Content of ward level data depends on numbers involved to ensure no disclosure.</p>

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
<p>Income Support: Annual scan of all claimants by DWP. Latest data available is for 2001 using 2003 ward boundaries on DWP website. Ward level data includes total number of claimants by gender, family type, whether working age or pensioner, other types of claimant, and age of dependents. Ward level data for 1998, 1999 and 2000 is available on ONS website in Neighbourhood Statistics using 1998 ward boundaries.</p>	Annual 100% scan of all claimants	Ward	✓	No current rurality indicator used but will use new definition if useful. Jobseekers allowance has replaced Income Support for unemployed people, but Income Support still paid to people not required to work – e.g. pensioners, lone parents and sick/disabled people. Numbers of claimants at ward level is sufficient not to be disclosive.
<p>Migrant Workers 2001/02: 100% scan of all workers flagged as migrant (arriving from outside the UK) on the National Insurance Recording System for the tax year 2001/02. Data includes information on migrant workers age, gender, region of UK residence, continent/country of origin.</p>	100% scan of workers flagged as migrant on National Insurance Recording System	County/ UA by request to DWP	<input type="checkbox"/>	No current rural/urban indicator used. Data broken down to GOR level as standard on website and figures are rounded to nearest 100. LA level data may be available by request from DWP as long as numbers involved are not disclosive. DWP do not hold data at lower levels than that. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas.
Other Socio-economic Data				
<p>Bereavement Benefit Statistics: Bi-annual 5% analysis of Widowed Parents Allowance and Bereavement Allowance by DWP. Data includes number of claimants by benefit category. Currently 100% data are only collected for new claims, although in future 100% data may be available with more detail.</p>	Bi-annual 5% scan of Pensions Strategy Computer System. 100% data on new claims	County/ UA by request to DWP	<input checked="" type="checkbox"/>	No current rural/urban breakdown. Theoretically 5% data are available at LA district level, but much of data may be suppressed at this level due to unreliability from large sampling errors. GOR level for 5% data are available. 100% data are available at LA level. Postcode information is held which could potentially be used to add new rural/urban definition, depending on usefulness and if numbers sufficient to ensure no disclosure.
<p>Housing Benefit & Council Tax Benefit: Data collected by DWP that includes: recipients of both benefits by tenure and whether in receipt of unemployment benefits; weekly amounts of benefit received.</p>	Quarterly 100% aggregate data collected from 2 statistical returns from LAs. Also annual 1% scan of claimants with more detailed data.	County/ UA	<input type="checkbox"/>	No current rural/urban breakdown. Quarterly data are aggregated to LA level. Breakdown into rural/urban areas may not be very sensitive at LA level as this covers large areas. The 1% annual data can only be broken down to GOR level due to unreliability of smaller numbers due to large sampling error. No postcode information is held for these benefit recipients and so this cannot be used to attach rural/urban indicators. Ward level data may be available in the future, but not currently.

Description of dataset	Collection type	Available spatial resolution	Potential for using new rurality definition	Comments
Housing Statistics: Comprehensive data on housing in England collected by the ODPM under the following topics: dwelling stock; house building; housing renewal; households, population estimates and projections; housing market; social housing; rents and tenancies; household characteristics; housing finance and household expenditure. The majority of data are only available down to LA level – however in future more data will be available down to ward level and will be available on the Neighbourhood Statistics website at ONS. The 2001 census data are also on the Neighbourhood Statistics website and it holds a lot of information on the type of housing, household characteristics and renting/tenure of the population at census night. The census data are available down to output area.	Data collected by ODPM regularly through a large series of statistical return forms, all detailed on the ODPM website. Large amount of housing and household information also collected from 2001 census.	Output area	✓	No current rural/urban indicator used within the datasets, and no immediate plans to use new rurality definition. 2001 census housing data are available down to census output area and this information can be linked with the other census data. ODPM does deposit limited data for selected housing statistics down to ward level in Neighbourhood Statistics website – however plans in future to extend the amount of data at ward level.
State Pension Statistics: Annual scan of all claimants by DWP. Latest data available is for 2001 using 2003 ward boundaries on DWP website. Ward level data includes total number of claimants by age and sex.	Annual 100% scan of all claimants	Ward	✓	No current rural indicator but will use new definition if useful. Numbers of claimants at ward level is sufficient not to be disclosive.

KEY

- ✓ Data can be disaggregated and published at areas smaller than LA/UA or PCT level.
- Data can be disaggregated to smaller areas than LA/UA and PCT for attachment of rural/urban indicators or for statistical modelling etc., but data cannot be published at these smaller areas.
- Data can only be disaggregated at the most to LA/UA or PCT level and rural/urban indicators cannot be attached at lower levels.
- ✗ Data are unlikely to be analysed meaningfully using rural/urban distinctions.

List of contributors

ORGANISATION	TITLE OF DATASET	CONTACT
Health Protection Agency	Communicable Diseases	Andrew Chronias
Health and Safety Executive	Fatal Injuries Survey	Ingrid Summersgull
University of Essex	British Household Panel Survey	Heather Laurie
Office of Deputy Prime Minister	Housing Statistics	Oscar Yau
Department of Education & Skills	Children adopted from Care, England	Anne Foulger
Department of Health	Guardianship under the Mental Health Act 1983	Kevin Downy
Department of Health	Community Care Statistics - Home Care Services for Adults	Kate Anderson
Dept for Works and Pensions	Widow's Benefit, attendance allowances etc.	Carol Jenkinson
Office for National Statistics	Births survey and Deaths survey (Census) Line 93 and 97	N O'Rourke
Defra	Agricultural Census	Miles Templeton
Office for National Statistics	Cancer Registry	Dr Mike Quinn
Department of Health	Health Survey for England	Patrick Tucker
CHI	Patient Survey	Dr Sarah Scobie
Department of Health	Hospital waiting times	Richard Winstanley
Department of Health	Breast Screening	Lesz Lancucki
Department of Health	NHS Hospital & Community Health Services non-medical staff in England	Stephen Triffitt
Department of Health	National Drug Treatment monitoring system (from 1 April 2001) (was Regional Drug misuse)	Rosemary Lane
Office for National Statistics	Contraception and Sexual Health	Dr H Meltzer
Office for National Statistics	Mental Health of Carers	Dr H Meltzer
Home Office	British Crime Survey	Anna Upson
Home Office	Probation Statistics	Gary Renshaw
Department of Health	Personal Social Services Expenditure	Ian Flaxton
National Primary Care Research and Development Centre	National primary Care Groups and Trusts Research Resource Database	Andrew Wagner
Department of Education & Skills	Children Accommodated in Secure Units, England and Wales	Anne Foulger

Discussion points

Preamble

This project (funded by Defra) aims to understand how reliable national datasets are in describing poor health in rural areas. The aim of the meetings is to explore this qualitatively with key people directly involved in the datasets of interest. The information from these meetings will be used to supplement a statistical review of selected datasets (being undertaken by the Centre for Medical Statistics at Keele University) in order to examine the 'rural relevance' of different datasets. The following discussion points should enable us to explore the rural relevance of the data and assess the ease or any potential difficulties in applying the new definition of urban/rural areas.

In the final report to Defra, the information from the meetings and the statistical review will be summarised and a list of general recommendations will be made about which datasets are most appropriate for further in-depth analysis from the rural health perspective.

Organisation:

Title of Dataset:

Contact:

Date:

Interviewee: PD LJ JAD

General points

1. What is the objective for collecting the dataset?
2. Who are the end users?
3. What population does the data cover? (eg geography, occupation, **age range (<16 or >75s?)** etc)
4. What is the sampling method?

Data quality

5. What do you think the data quality is of these/this dataset?

6. Are there any data quality checking systems in place and if so what are these? (built in intelligence? Manual or electronic data checking systems?)

Confidentiality

7. How are 'small numbers' dealt with in the datasets (rounded down to zero? /blurred?)?

Rural relevance

8. Does your organisation have a specific remit to gather data on rural areas?
9. Does the data cover rural areas?
10. Does your organisation use a definition of rurality? If so, what is it and is the data categorised by urban/rural?
11. Are you aware of the new classification of urban/rural areas based on Census Output Areas from the Office for National Statistics?
(The project is being led by the Office for National Statistics (ONS) and is being co-sponsored by the ONS, the Office of the Deputy Prime Minister (ODPM), the Department for Environment, Food and Rural Affairs (Defra), the Countryside Agency (CA) and the Welsh Assembly Government. The work is being undertaken by a consortium led by Birkbeck College and involving the University of Sheffield, the University of Glamorgan, and Geowise)
12. If so, do you have any plans in place to analyse your dataset using this classification?
13. What is the smallest geographical unit at which the data is available?
14. Do you feel that the data are representative of rural areas?
15. Do the data identify urban-rural differences?
16. Do the data identify intra-rural variations?
17. Are there other data sets to which this data set could be linked easily to facilitate analysis of rural health issues?
18. Any other issues that haven't been covered that you would like to raise?

Closing points

Thank you for your contribution. With your agreement we would like to include you in the list of contributors to this research.

Organisation Address	Phone E-mail Website	Nature of research/organisation
Department of Health (Legislation and Regulation Team)	020 7210 5448 Alexander.Wallis@doh.gsi.gov.uk http://www.doh.gov.uk/	Is aware that Defra are planning a customer research project that touches on health issues in rural areas (perhaps therefore 'small area' data sets). Info on this is as follows: The objective of the research is to interview two key customer groups, - people who live in urban/rural areas and managers of businesses operating in urban/rural areas - to obtain baseline information.
Rural Stress Support Network (West Midlands Region) N.C.C.C. Upper Bar Newport Shropshire TF10 7EH	0870 707 4445 info@ruralstresswestmids.org.uk http://www.ruralstresswestmids.org.uk	Has recently conducted a snap-shot survey on a sample of their cases. The organisation collects and analyses relevant information appropriately and systematically uses it to help managers and staff to evaluate and improve the organisation and its activities.
Rural Mental Health Research Group West Hampshire NHS Trust Maples Tatchbury Mount Calmore Southampton SO40 2RZ	023 8087 4330 Andrew.Mayers@WHT.NHS.UK	Research on access to, and provision of, mental health services in rural areas, as part of the National Mental Health Partnership. This could include issues such as the provision of dedicated staff to remote areas, alternative delivery systems, the mental health of agricultural workers, etc.

<p>MARES Old Amersham Farm High Street Old Amersham Bucks Buckinghamshire HP7 0EJ</p>	<p>01494 729522 skrokka@yahoo.com http://www.mares.org.uk</p>	<p>Provide therapy for people with some form of mental health difficulty in a rural setting using agri-equestrian activities especially, for those from black and ethnic minority communities. About to carry out a piece of research to understand why the local voluntary and charity organisations have not been able to attract volunteers from the local black and ethnic minority communities.</p>
<p>Derbyshire Rural Community Council Church Street Wirksworth Matlock Derbyshire DE4 4EY</p>	<p>01629 821923 graham@derbysrcc.org.uk http://www.derbysrcc.org.uk/</p>	<p>* 3-year research into effectiveness of a Melody, Memory and Movement Community-Funded programme in 15 rural residential care homes in Derbyshire Dales.</p> <p>* Small scale research on the health and social care benefits of a Pottery Paint and Petals programme in 8 residential care homes.</p>
<p>Harper Adams University College Edgmond Newport Shropshire TF10 8NB</p>	<p>01952 815330 care@harper-adams.ac.uk http://www.careinthecountryside.net/</p>	<p>CaRE (Convalescent and Recuperative Environment) Project is currently looking into the need for more choice in the countryside in terms of short breaks for carers or for those with special or particular needs possibly to be funded by the Social Services direct payments scheme. Currently collating information and hoping to go forward in the New Year with some input and advice from the Health Services Management Centre at the University of Birmingham.</p>

<p>Redcar & Cleveland Mind Dove House 5 Turner Street Redcar TS10 1AY</p>	<p>01642 296052 info@randcmind.org http://www.goodmindguide.org/about_randc.htm</p>	<p>Outreach day services operating daily on a rota around the rural areas, centred on local transport hubs; in small market towns and villages. Clientelle are people in the community with a mental health need/issue operating on an open access drop-in model. Regularly monitor information requests and have statistics on how many people enquired, how many enquiries and what subject (in broad terms) they enquired about. Have also undertaken a small sample questionnaire on stress with the general public in the area as part of our events promoting World Mental Health Day.</p>
<p>Clinical Supervision</p>	<p>www.clinical-supervision.com</p>	<p>Developing a website to deliver psychotherapy for common mental health problems to the people of Conwy, Denbighshire and Wrexham. Website: http://outreach-online.com</p>
<p>NHS Health Scotland Clifton House Clifton Place Glasgow G3 7LS</p>	<p>0141 300 1016 bruce.whyte@phis.csa.scot.nhs.uk</p>	<p>Three projects. One to update existing constituency profiles for Scottish MSPs Constituencies. Second = project to produce community health and well being profiles in 70 areas across Scotland. This project aims to produce a range of indicators down to postcode level for community areas roughly corresponding to the pre-existing Local Healthcare Co-operative areas and emerging Community Health Partnership areas - these are primary care planning groups in use in Scotland. Third project is a pilot to record primary care patient contacts in remote rural areas of Scotland. A health and well being survey has been undertaken as part of this project (contact = Jane Farmer at Aberdeen University j.farmer@abdn.ac.uk)</p>

<p>Berwick Family Centre 6a Grove Gardens Tweedmouth Berwick on Tweed TD15 2EN</p>	<p>01289 308822 berfamc@aol.com</p>	<p>Berwick Family Centre is a part of Berwick Borough Local Strategic Partnership, which has a sector board dedicated to Health and Well being. This Sector Board is made up of individuals and organisations that are interested in health in its widest sense. The Board will be looking at issues that effect people living in rural areas of the Borough.</p>
<p>Northumberland Care Trust</p>	<p>shona.haining@northumberlandca retrust.nus.uk</p>	<p>Rural issues are one of their key themes. May be interested in collaborating in the future.</p>
<p>Craven, Harrogate & Rural District NHS PCT</p>	<p>01423 859610 lea.marshall@chrd-pct.nhs.uk</p>	<p>Has carried out analysis work on Emergency Admissions from Craven and Harrogate, which led to a strategy for emergency care.</p>
<p>Carlisle & District NHS PCT Wavell Drive Rosehill Carlisle CA1 2SE</p>	<p>01228 603500 catherine.gregson@ncumbria.nhs. uk</p>	<p>Under the lead of Dr Jean Vickers, Director of Public Health for West Cumbria PCT, they are conducting a health survey of 1 in 10 of their adult population (a sample size chosen to take account of the sparse nature of some of their communities. Anticipate results being available in Spring 2004. They have also carried out analyses at ward level of routinely available data.</p>
<p>Shropshire & Wrekin PCT</p>	<p>01952 226013</p>	<p>Concerned about access to services in rural Shropshire, so have begun carrying out hearing tests for new born babies on first home visit by health visitor, rather than in hospital. New research starting in December 2003 for 12 months looking into effectiveness of this new service. Menna Reese is actually carrying out the research. Research Head, Dr Chris Bunting (Paediatric Audio Specialit, Shropshire) - chris.bunting@shropcomm.wmids.nhs.uk 01743 450825</p>