

Rural markers in datasets held at the UK Data Archive

UK Data Archive

Author: Veerle Van den Eynden

Date: 30 September 2009

The UK Data Archive (UKDA) holds a collection of 4650 studies or datasets, covering a range of economic, social and historical data. Datasets result from research and surveys carried out by government departments, public authorities, research institutes, private companies and academic researchers. UKDA acquires, preserves and enhances datasets and disseminates them freely online for research and educational purposes.

A review was carried out of the UKDA data catalogue in order to identify national scale survey data for England that have potential to be used for rural and urban analyses by the Commission for Rural Communities.

To this extend, two reviews were carried out. In a first instance datasets were identified that contain rural variables. The datasets identified fit the following criteria:

- contain a rural / urban variable (e.g. ONS 2004 rural / urban area classification) or contain Output Areas as spatial units
- recent datasets (post 2000), except in the case of time series and longitudinal data collections where entire series are included
- national datasets for England
- with representative population sample.

Secondly, all recent national survey datasets for England (post 2000), with representative population sample that do not contain a rural / urban variable were identified and data owners contacted to investigate:

- whether a rural / urban variable may exist for the dataset and can be made available
- whether the dataset can be retrospectively coded to include a rural / urban variable, based on the presence of postcodes in the dataset.

Review procedures

The review consisted of a series of queries and searches of the entire UKDA catalogue to identify national datasets post 2000, with and without rural variables.

For each identified dataset:

- data files were checked for presence of the rural variables and to verify variable values and position of the variable in the data file
- related studies were investigated for presence / absence of similar variables
- study documentation was checked for information on rural / urban variables
- data processing notes were reviewed for any variable alteration or removal due to confidentiality
- access conditions of the dataset were verified

For national datasets without rural variable, data owners were contacted by email to seek information on the possible availability of rural markers or postcodes and the option of recoding for rural markers based on postcodes.

Results

All datasets that fit the criteria are presented in an attached Excel table named 'UKDA-RuralMarkersReviewResults_30-09-09', which contains two worksheets: 'DatasetsWithRuralVariables' and 'DatasetsNoRuralVariables'.

All identified datasets are quantitative data which are accessible in SPSS format. No national qualitative datasets with representative samples were found to exist.

Each worksheet contains the following fields for the datasets:

UKDA study number	The study number of the dataset in the UKDA data catalogue
Dataset title	
Depositor	Data owner
Temporal coverage	Reference year(s) for the dataset
Geographic coverage	For national datasets this may be England, England and Wales, Great Britain or United Kingdom
Geographic variables / spatial units	The smallest spatial units for the dataset
Sample size	Sample size for the dataset (for some with indication of sample size for England only)
Rural marker ID	Variable name of the rural variable in the dataset
Rural marker description	
Rural marker values	Different values distinguished for the variable
Rural marker position in data file	The position of the identified variable within the data file
Access conditions for dataset	
URL to UKDA catalogue record	
Availability of variable or postcode	Whether the variable is available or not within the dataset held at UKDA, or can be obtained from the data owner; with conditional details where applicable. Or whether postcodes exist and retrospective coding can be carried out.
Additional notes	

Datasets with rural variables

A total of 49 datasets or dataset series (time-series or longitudinal studies) have been identified that contain a rural variable. Some contain various variables.

A range of different types of rural markers were found to be used in datasets, with the ONS 2004 rural / urban definition dominating datasets since 2004.

Rural marker type	Number of datasets where variable type used
ONS 2004 rural / urban definition, distinguishing either 8 or 6 categories	23 datasets
Binary rural / urban variable	6 datasets
Variable distinguishing rural / urban / other category	2 datasets
Other variable indicating 'rural areas'	8 datasets
Respondent evaluation of rural / urban area	7 datasets
Indirect indication through settlement size or population density	3 datasets

Sixteen datasets or dataset series that contain the ONS 2004 rural / urban definition are available from UKDA, either through standard UKDA access or via Special License or Approved Researcher access:

- British Crime Survey 2003-2008
- British Household Panel Survey 1991-2008
- English House Condition Survey 2001-2006
- Expenditure and Food Survey 2004
- Experience and Expression in the Fear of Crime 2003-2004
- Health Survey for England 2003-2007
- Home Office Citizenship Survey 2005
- Longitudinal Study of Young People in England 2004-2006
- Mental Health of Children and Young People in Great Britain 2004
- Millennium Cohort Study 2001-2006
- National Survey of Volunteering and Charitable Giving 2006-2007
- Offending, Crime and Justice Survey 2003-2005
- Quarterly Labour Force Survey 2005-2009
- Survey of Public Attitudes and Behaviours toward the Environment 2007
- Taking Part: the National Survey of Culture, Leisure and Sport 2005-2006
- Workplace Employee Relations Survey 2004

For seven datasets held at UKDA, the ONS 2004 rural / urban definition can be requested from the data owner. The variable has either been removed by the data owner due to disclosure risk, or has been created at a later stage and not been deposited at UKDA:

- 1970 British Cohort Study 2004, 2008
- Annual population survey 2004-2008
- English Longitudinal Study of Ageing
- Family Resources Survey 2004-2008
- National Child Development Study 2004
- ONS Omnibus Survey (Opinions survey)
- Youth Cohort Study

Various data owners, especially government departments, consider the ONS 2004 rural / urban definition to pose a risk of disclosure, as the variable is based directly on postcodes. They may therefore remove the variable from datasets offered to UKDA.

Datasets dating from before 2004 do not contain the ONS 2004 rural / urban definition, but may contain a variety of other rural / urban variables, either simply distinguishing rural from urban areas; or containing a range of values, some of which indicate rural areas. Details of the categorical values used for each type of variable are included in the results table.

For time series, consistency in the use of rural / urban variables throughout the consecutive surveys does not always exist. Variables used may have changed through time.

Datasets without rural variables

Nineteen national datasets without rural variables were identified as fitting the criteria and the respective data owners contacted to assess whether postcodes exist for the datasets that would enable retrospective coding for rural / urban variables. For various datasets the data owners are willing to carry out recoding. Data owners are usually not willing to provide postcodes, due to the disclosure risk and the confidentiality agreements drawn up with respondents. Details are included in the results table.

Access to datasets

URL links included in the results file lead directly to the UKDA catalogue record for each dataset (or for a group of datasets in the case of time-series or longitudinal studies). This catalogue record is in the public domain and thus accessible to all.

Access conditions for datasets and for rural / urban variables are indicated in the results file.

Most datasets are available via the UKDA standard conditions of use of data. This requires users to register with UKDA and to sign an End Use License before datasets can be downloaded. The End Use License specifies conditions of data use.

For some datasets and to access some rural / urban variables Special License conditions exist. Users may either be required to apply for ONS Approved Researcher status; may need to apply to UKDA for a Special License or specific confidentiality conditions may be imposed by data owners. Details of the conditions can be found in the catalogue record for the datasets, or are detailed in the results tables.

A summary of special access conditions that exists for major surveys held at UKDA can be found at:

<http://www.data-archive.ac.uk/orderingData/termsandconditions.asp>

Review results can be found in accompanying MS Excel file 'UKDA-RuralMarkersReviewResults_30-09-09'.

Contact

For further information about this review, please contact:

Veerle Van den Eynden
Rural Economy and Land Use programme - Data Support Service
UK Data Archive
University of Essex
Wivenhoe Park

Colchester CO4 3SQ
Tel: +44 (0) 1206 872234
veerle@essex.ac.uk