

# Neighbourhood Data to be used with the 2005 British Election Study

Ron Johnston and Rich Harris

*School of Geographical Sciences, University of Bristol*

Much has been written about the likelihood that there is a 'neighbourhood effect' in voting at British general elections, whereby some individuals at least are influenced by their local context to vote in ways that might not have been expected given their personal characteristics. This hypothesis has remained, at best, only partly verified-falsified because of (a) the paucity of data and (b) the lack of any empirical appreciation of the processes involved.

Plenty of data analyses have been conducted whose findings are consistent with the hypothesis, but almost all of these have been at spatial scales (usually that of the constituency) which are incommensurate with the local milieux in which neighbourhood effects are assumed to operate. Recently, however, work using data taken from the 1997 British Election Study and the 1991-2001 British Household Panel Study has provided much more convincing evidence of patterns of voting consistent with the neighbourhood effect at relevant spatial scales. These used the concept of a 'bespoke neighbourhood' for each individual respondent to those two surveys; combining information on the respondent's home location (i.e. their postcode) with small-area census data, indicators of the characteristics of each person's neighbourhood (down to the nearest 500 persons to their home and all persons living within 250 metres of that location) were used to allow tests of the hypothesis. These showed very clear patterns whereby similar people (defined by, for example, their socio-economic class and their housing tenure) voted differently in different types of neighbourhood. (Examples of this work can be found in MacAllister et al, 2001, and Johnston et al, 2005.)

Given the results of these analyses, it was decided to try and produce neighbourhood data for the 2005 BES, using the 2001 census. This used a different architecture for reporting results at small spatial scales from previous censuses, and it was decided that rather than create a wide range of bespoke neighbourhoods for each respondent, data sets would be produced using the reporting units adopted by the census. *Unfortunately, because of a different architecture of reporting units in Scotland, it has only proved possible to do this for England and Wales.*

With the cooperation of the two bodies which undertook the survey work for the 2005 BES – the National Centre for Social Research and YouGov – we were provided with the postcodes for all individuals sampled in the two surveys (both of which had a pre- and post-election wave; the YouGov survey also included a rolling campaign panel). These were used to place each individual in the following areas, defined by the Office of National Statistics.

Area Type	Number	Total Population 2001	
		Mean	SD
Output Area (OA)	175,434	297	72
Lower Layer Super OA	34,378	1,513	203
Middle Layer Super OA	7,194	7,234	1,320
Census Ward 2001	9,278	5,604	4037
Electoral Ward 2004	8,800	5,909	4,066
Parliamentary Constituency	569	91,387	10,872
Urban Settlement	4,726		

Output Areas (OAs) are the basic reporting units for 2001 census data, and are significantly smaller than the enumeration districts used for 1991 and previous censuses. They were defined as amalgamations of postcodes so as to create small areas with considerable internal homogeneity using two variables – dwelling type and dwelling tenure – with constraints for size and shape; the minimum size was 40 households, but the recommended average was 125 households. In addition, they had to nest within the wards in existence at the time the data were tabulated in 2003 (the Census Wards in the above table). The data were later reorganised to fit within the wards extant in 2004 (the Electoral Wards in the above table) following the continuing work of the Local Government Committee for England (the former Local Government Boundary Commission). (On the construction of Output Areas, see [http://www.statistics.gov.uk/geography/census\\_geog.asp#oa](http://www.statistics.gov.uk/geography/census_geog.asp#oa); see also <http://www.statistics.gov.uk/geography/downloads/georoadshowpaper.pdf>.)

The OAs also nest within the Lower Layer and Middle Layer Super OAs. The Lower Layer Super OAs (LLSOAs) were designed as groups of typically 4-6 OAs, with a minimum population of 1000 and a design mean of c.1500: they were also to nest within the 2001 Census Wards. The Middle Layer Super OAs (MLSOAs) were designed as groups of LLSOAs having a minimum population of 5,000 and a mean of 7,200; they had to nest within Local Authorities. (For full details on the creation of SOAs see <http://www.statistics.gov.uk/geography/soa.asp#3layers>). LLSOAs and MLSOAs are in the size range of wards but, as the data in the above table indicate, vary much less in their populations: some wards are very large (populations c.20,000, as in Birmingham) whereas others are small, notably in rural areas. (There is also an intention to create Upper Layer SOAs, with a minimum size of 25,000, but these have yet to be defined and, by their size, would be much too large to be of significant value in neighbourhood effect studies.)

The Parliamentary Constituencies are those used in the 2005 general election. The Urban Places were defined by ONS through an analysis of built-up areas. All OAs were classified as either urban or rural and contiguous urban OAs were combined to form the defined urban places. Within the conurbations and larger urban areas separate settlements were also defined – e.g. Keighley was separately identified from Bradford within Bradford Metropolitan District. Each BES respondent was located within her/his urban place; a number who lived in rural areas (i.e. outside the defined places) have no data in this set.

To produce the BES neighbourhood data, census data were obtained from the relevant websites and disks for a wide range of variables considered relevant to the study of political attitudes and behaviours. From these, and after consultation with potential users, a set of 61 derived variables was decided upon: all but one are percentages of a given denominator – the exception is the population density measure (persons per hectare). These were calculated for each of the areas at each of the seven scales, and the relevant value was then extracted for each BES respondent. The variables are shown in the Appendix at the end of this paper.

There are thus seven data sets for each of the two BES surveys:

1. prepost: the face-to-face surveys conducted pre-campaign and after the election, containing 3578 observations (in England and Wales); and
2. yougov: the internet surveys conducted at the beginning of the campaign and after the election, plus the rolling campaign panel, containing 7113 observations (in England and Wales).

Each comprises:

- The serial number for the relevant survey, allowing ease of file merger with the survey data;
- The 61 derived variables for which the short titles begin with a letter indicating which data set: o – OA; l – LLSOA; m – MLSOA; x – census ward; w – electoral ward; c – parliamentary constituency; u – urban area; and
- The area (in hectares) and population of that unit.

In addition, two of the data sets have additional information:

- The OA data sets include information derived from a recently-published classification of all Output Areas produced at the University of Leeds for ONS. This hierarchical classification has three levels, comprising 7, 21 and 52 nested groups: the first two levels have names, and these are included with the group numbering on the data sets (full details of the classification can be found at <http://www.geog.leeds.ac.uk/people/d.vickers/OAclassinfo.html>; and
- The constituency data sets include information on the results in the constituencies at each of the last four general elections – 2005, 2001, 1997 and 1992: those for 1992 are estimates for the new constituencies – taken from Pippa Norris’s British Parliamentary Constituency Database 1992-2005: <http://ksghome.harvard.edu/~pnorris/Data/Data.htm>.

The datasets (spss .sav files) are:

<b>prepost</b>	<b>yougov</b>
ppOA	ygOA
ppLSOA	ygLSOA
ppMSOA	ygMSOA
ppCWARD	ygCWARD
ppEWARD	ygEWARD
ppCON	ygCON
ppURBAN	ygURBAN

On none of these files, other than that for the parliamentary constituencies, is it possible to identify which area (OA, LLSOA etc.) the respondent lived in. It was

central to this project – as in the earlier bespoke neighbourhood studies – that the respondents’ confidentiality be totally protected.

Note that with all census data, ONS took steps to protect respondent confidentiality by the process known as ‘Barnardisation’ (after George Barnard, founding Professor of Statistics at Essex). Where the number of observations in a cell is small, then a random number in the range +/-3 was added to it. In a very small number of cases, especially with OAs, this may lead to odd values for some derived variables.

## **Scotland**

Unfortunately, although the General Register Office for Scotland created Output Areas in the same way as ONS did for England and Wales (and Northern Ireland) it did not amalgamate them into LLSOAs, MLSOAs etc. Hence we have been unable to create the neighbourhood statistics files for the whole of Great Britain. However, the National Classification of Output Areas covers Scotland and we have used this to create some data for the Scottish samples. For all sampled individuals in both the prepost and YouGov surveys, therefore, we have produced a file placing each respondent in her/his category, at each of the three levels. These files are

ppSCOTOA

ygSCOTOA

## **Additional data**

The 61 derived variables provided here far from exhaust the potential available from the census data. The core files with all of the census data used here, from which the derived variables were obtained, are held by Ron Johnston, and it may be that if individual researchers want other variables to be calculated for one or both of the surveys at particular scales this can be done – within the constraints of time and data availability.

Ron has also undertaken factor analyses using 34 of the derived variables at six of the scales: a factor analysis on the urban area data was not done because it included places varying in size from under 500 to over 8 million. Those factor analyses were conducted on the full set of areas, not just those which included survey respondents – i.e. all 175,434 OAs etc. He has files of the factor scores, and should individuals wish access to these, again they should contact him.

Finally, the National Centre provided postcodes for the full sample, and not just those who responded. If anybody wishes to analyse the geography of the sample, rather than that of the respondents, they should contact Ron. (The sample is split into five categories: those who agreed an interview; those who could not be contacted; those who refused; others; and those addresses that were out of scope.)

Ron can be contacted at [R.Johnston@bristol.ac.uk](mailto:R.Johnston@bristol.ac.uk)

## **Acknowledgements**

In producing these data sets we are grateful to the BES Principal Investigators (Harold Clarke, David Sanders, Marianne Stewart and Paul Whiteley) for agreeing to our proposal to create these datasets, and David Sanders for finding some money from within his budget; to Katarina Thomson of the National Centre and Joe Twyman of YouGov for their ready cooperation in providing the needed postcodes; to Dan Vickers for agreeing that we could include the data from the National Classification of Output Areas; and Pippa Norris for agreeing that we could include data from the British Parliamentary Constituency Database 1992-2005. Above all, however, we are grateful to Alison Smith who did all the work involved in creating the data sets from the postcode and census files: she did a marvellous job.

## **References**

- Johnston, R. J., Propper, C., Burgess, S., Sarker, R., Bolster, A. and Jones, K. 2005: Spatial scale and the neighborhood effect: multinomial models of voting at two recent British general elections. *British Journal of Political Science* 35:
- MacAllister, I., Johnston, R. J., Pattie, C. J., Tunstall, H., Dorling, D. F. L. and Rossiter, D. J. (2001) Class dealignment and the neighbourhood effect: Miller revisited. *British Journal of Political Science* 31: 41-60.

## Appendix

### The 61 derived variables in the datasets.

Name	Variable
popden	population density
15under	% 15 and under
firstv	% first-time voters 2005
elect	% 18+ 2005
nochild	% no-child households
bornouk	% born outside UK
black	% black
sasian	% south asian
nonwh	% non-white
nonchr	% non-christian
muslim	% muslim
ngheal	% not in good health
ltermi	% long-term illness
noqual	% adults no qualifications
hiqua	% adults level 4-5 qualifications
ptime	% employees working part-time
ftstuds	% full-time students
persick	% economically inactive sick/disabled
profman	% economically active professional/managerial
skilled	% economically active skilled trades
element	% economically active elementary jobs
pbour	% economically active petit bourgeoisie
routine	% economically active routine/semi-routine occupations
agric	% employed agriculture
manuf	% employed manufacture
finance	% employed finance
detach	% population in detached homes
house	% population living in house or bungalow
ownout	% population on owned outright homes
owned	% population in owned homes (inc mortgage)
socrent	% population in socially-rented homes
onepp	% households one-person pensioner
onenp	% households one-person non-pensioner
cohabit	% households co-habiting couples
hochil	% households with children
lonechi	% households lone-parent with children
studhou	% households multi-occupancy students
othmult	% households other multi-occupancy
1perroo	% households living one per room or less
notch	% households no central heating

nocar	% households no car
3plcar	% households 3 or more cars
migrate	% people migrated in last year
outuk	% who migrated from outside UK last year
nwmig	% non-whites migrated in last year
unemp	% economically active unemployed
rose1	% adults new socio-economic classification level 1
rose2	% adults new socio-economic classification level 2
rose3	% adults new socio-economic classification level 3
rose4	% adults new socio-economic classification level 4
rose5	% adults new socio-economic classification level 5
rose6	% adults new socio-economic classification level 6
7rooms	% households with 7 or more rooms
hipay	% economically active in high-paid private sector occupations
wales	% population born in Wales
ftime	% economically active working full-time
65plus	% aged 65 and over
75plus	% aged 75 and over
council	% population in homes rented from Council
retired	% population who are retired
stud	% population who are students