



Consultation document

The 2011 Census:
Initial View on Content for England and Wales
– Derived Variables

Derived Variables for the 2011 Census

1. Introduction

Initial views on the possible content of the 2011 Census for England and Wales are outlined in the 2011 Census consultation document, available on the National Statistics web-site:

http://www.statistics.gov.uk/about/consultations/downloads/2011Census_consultation_content.pdf

In keeping with the general shift in ONS to a theme based approach to information provision, users are being asked to respond to the above consultation document on a topic by topic basis. It is important however, that this approach also captures cross-topic user information requirements.

This paper supports the consultation document by providing further information on issues for consideration in the provision of cross-topic information and the derivation of variables from 2011 Census data.

2. Derived variables (DVs)

The term "derived variable" refers to variables that are constructed by combining individual data items. For example in 2001 the derived variable "living arrangements" was derived by combining the responses to the question on legal marital status and the responses to the relationship question.

The potential for derived variables is therefore wholly dependent on the topics and questions that are included in the census. It is recognised, however, that this presents a "chicken and egg" situation where user need for derived variables could drive question inclusion or vice versa. It is therefore essential to gain an understanding of user needs for derived variables at an early stage of census questionnaire development.

3. Output of Standard Derived Variables

In designing derived variables for standard output from the 2011 Census, account will be taken of the need for comparability with 2001 Census standard derived variables. However, other factors are likely to necessitate some changes to output. Inclusion of any new topics in the 2011 Census would give scope for new variables to be derived. Conversely any omissions of topics previously included in the census may limit comparability with previous derived variable output.

For standard and key classifications the need for harmonisation with other government surveys both in terms of question format and in terms of outputs will be taken into account. Inputs and outputs of the three censuses taking place across the U.K will also be harmonised as far as possible.

Socio-demographic changes such as the U.K ageing population and the increase in the proportion of children living in lone-parent families will also be taken into account in designing variables to be derived as standard output.

A paper discussing these changes is available on the National Statistics web-site: http://www.statistics.gov.uk/about/consultations/downloads/2011Census_consultation_socio_demographic_context.pdf

Although the above factors will influence the design of variables to be derived as standard, the key determinant will be current and future user requirements.

4. Standard Classifications and Key Derived Variables

The following standard classifications and key derived variables have been identified and have fed into requirements for data as outlined in section 3 of the Census consultation document.

- National Statistics Socio-economic Classification (NS-SEC)
- Household and Family Structures
- Measures of Deprivation

NS-SEC

NS-SEC is the ONS standard socio-economic classification used in all official statistics and surveys. The derivation of NS-SEC from the Census supports the case for the inclusion of questions on occupation, industry, qualifications, activity last week, time since last worked, employee/self-employed, supervisor, and student in full-time education. The three pieces of information used to derive NS-SEC in 2001 were:

- the unit group code of SOC 2000
- employment status
- size of organisation

A SOC 2000 unit group occupation code was assigned using the questions asking 'what is the full title of your main job?' and 'describe what you do in your main job'. The qualifications question was used to assist with coding to professional occupations.

The following questions, along with the SOC 2000 code, were used to derive the employment status/size of organisation codes:

- 'do you work as an employee or are you self-employed?'
- 'do you supervise any other employees?'
- 'how many people work for your employer at the place where you work?'

There is strong user demand for the collection of the necessary data to enable NS-SEC to be derived from the 2011 Census. However, as detailed above, the full version of NS-SEC requires the inclusion of multiple data items, some of which respondents found difficult to answer in the 2001 Census. This necessitated a relatively high level of imputation for some of these data items, (in particular workplace size), in the derivation of NS-SEC.

The evaluation report '*Census 2001 Review and Evaluation: Edit and Imputation*' is available on the National Statistics web-site:

http://www.statistics.gov.uk/census2001/proj_eai.asp

These factors, together with constraints on physical space in the questionnaire mean that consideration will be given to the exclusion and/or simplification of some labour market and employment questions. It may not be possible therefore to derive the full NS-SEC classification. However it is possible to derive a reduced version of the NS-SEC classification that does not require detailed employment status information.

Household structure and Family structure

Information about the social composition of households and families is a key requirement of census users. In the 2001 Census questions on relationships between household members were not fully completed by all respondents resulting in some difficulties in derivation of household and family structure variables. As a consequence the relationship matrix question format used to obtain this information in the 2001 Census will be reviewed. The exact form of household and family structure information derived from the 2011 Census may therefore vary from that available from 2001.

Measures of Deprivation

A number of measures of household deprivation (employment, education, health and disability, housing) were derived as standard output from the 2001 Census. Although there has been some criticism of the exact form of some of these, there is continuing user demand for deprivation measures at low geographical levels.

There is a possibility that a question relating to income will be included in the 2011 Census. It is anticipated that any Census outputs on income will be broad-banded classificatory variables. User requirements for indirect measures of deprivation will be assessed in relation to the possible inclusion of the more direct deprivation measure of income band.

In addition to household deprivation measures derived solely from census data, 2001 Census data were also a key input for the creation of the Indices of Deprivation 2004 (ID2004). ID2004 were commissioned by the Office for the Deputy Prime Minister (ODPM) and created by the Social Disadvantage Research Centre at Oxford University. ID2004 relate to England only. Spatial measures of deprivation, similar to those published by ODPM and including limited amounts of census data, have been published for Northern Ireland, Scotland and Wales.

ID2004 play a key role in identifying the most disadvantaged areas in the country and there is therefore a high user demand. In the main the census topics that were utilised were population and residence variables but additionally, measures of household over-crowding and the central heating question were also used.

5. User requirements for standard derived variables -Responding to the consultation document

User requirements for standard derived variables from the 2011 Census should be included in the response to the Census consultation document on the possible content of the 2011 Census for England and Wales. The response questionnaire is available at on the National Statistics web-site:

http://www.statistics.gov.uk/about/consultations/downloads/2011Census_consultation_response_template.doc

The response should consider the likely relevance of any identified derived variables in 2011 and beyond as well as derived variables likely to become important in the future and/or those with specific policy implications. The potential for new derived variables based on the possible inclusion of new topics, discussed in section 3 of the Census consultation document, should also be taken into consideration in the response.

6. Non-standard Derived Variables

In addition to the output of standard derived variables, it was possible to commission user specified derived variables from the 2001 Census. Although output strategies for the 2011 Census have not been fully considered it is anticipated that the derivation of user defined derived variables will be possible from the 2011 Census.

The Market Research Society developed a method for using 2001 Census occupation and employment related information to provide an approximation of Social Grade, the socio-economic classification used by these industries. Derivation of Social Grade from 2011 Census information will be dependent on the collection of the data items used in the algorithm to produce this variable. All of these data items are classified as items for which either it is considered there is a clear case for inclusion or further work will be under-taken prior to a decision being made about inclusion.

A Sample of Anonymised Records (SAR) has also recently been made available to registered users enabling user defined analysis of 2001 Census data. It is anticipated that non-disclosive microdata will be made available from the 2011 Census.

All standard and non-standard derived variable outputs from the 2011 Census will be subject to disclosure control and confidentiality methodology.

Angele Storey
