

## WERS 2004 OCCUPATIONAL DATA

### NOTE FOR USERS

The file 'Occupational data.sav' contains information on the composition of occupations for employees in Great Britain at the unit group (four-digit) level, based on the 2000 Standard Occupational Classification (SOC 2000), which can be matched to the WERS 2004 Survey of Employees. The file provides the following items for each occupational group:

- Gender composition;
- Proportion under the age of 22;
- Proportion aged 50 or more;
- Ethnic composition;
- Proportion with a long-term work-limiting disability;
- Average gross hourly wages;
- Proportion with highest educational qualification at each NVQ-equivalent level;
- National Statistics Socio-Economic Classification;
- Female and male CAMSIS scores.

The source of each of these variables is described in detail below. In order to match these items on to the WERS 2004 Survey of Employees (SEQ) it is necessary to ensure that the SEQ is sorted by serno and persid. The occupational data can then be matched to the SEQ using the following commands:

```
get file='XS04_SEQv2.sav'.
```

```
sort cases by serno persid.
```

```
match files file=*/table='Occupational data.sav'/by serno persid.
```

```
execute.
```

### **Derived variables**

#### Gender composition (occgen)

The 2004 Labour Force Survey (LFS) was used to calculate the gender composition of every four-digit occupation. The resulting variable (occgen) records the proportion of employees in the occupational group who are female.

#### Proportion under the age of 22 (occsb22)

The variable occsb22 records the proportion of employees with each unit group level occupation under the age of 22 according to the 2004 LFS.

#### Proportion aged 50 or more (occ50p)

The proportion of employees within each four-digit occupation aged 50 or more is recorded in occ50p and was derived from the 2004 LFS.

### Ethnic composition (occeth)

The ethnic composition of each four-digit occupation was calculated from the 2004 LFS and the resulting variable (occeth) contains the proportion of the occupation who were not white.

### Proportion with a long-term work-limiting disability (occdisab)

The 2004 LFS established whether respondents had a health problem or disability expected to last for more than one year, and if so, whether this problem affected the kind of paid work that they might do. From this, a variable, occdisab, was derived to indicate the proportion of each four-digit occupational group with a long-term work-limiting disability.

### Average gross hourly wages (occhrw)

The variable occhrw gives the gross average hourly wage rate, including overtime in the employee's main job, for each unit-group occupation, derived from the 2004 LFS. The hourly wage rate was truncated at £100 an hour, to avoid the average for any individual occupation being affected by a small number of outliers which may have been due to mis-recording of the wage or hours data, or wage rates which were exceptional for the occupation. Information on less than 0.1 per cent of employees was discarded as a result of this truncation. The lowest average gross hourly wage rate for any occupation was £4.95, whilst the highest was £33.78.

### Proportion with highest qualification at each NVQ-equivalent level

The 2004 LFS contained a variable which recorded the highest educational qualification held by individuals compared against National Vocational Qualifications. This was itself derived from more detailed questions on the academic and vocational credentials that they held. For each four-digit occupation, the proportion of employees whose highest qualification fell into each NVQ-equivalent level was calculated. For example, 19 per cent of call centre agents held the equivalent of an NVQ level 4 or higher. A further 24 per cent had the equivalent of an NVQ level 3, 3 per cent had completed a trade apprenticeship, 28 per cent had the equivalent of an NVQ level 2, 16 per cent had a qualification below NVQ level 2, 6 per cent had a qualification that could not be categorised and 5 per cent had no qualifications at all.

### National Statistics Socio-Economic Classification (ns\_sec)

The NS-SEC is determined by the employment status of the individual, as well as their occupation. The document 'The National Statistics Socio-economic Classification: User Manual' gives more details of how the NS-SEC is determined. This can be found at: [http://www.statistics.gov.uk/methods\\_quality/ns\\_sec/downloads/NS-SEC\\_User.pdf](http://www.statistics.gov.uk/methods_quality/ns_sec/downloads/NS-SEC_User.pdf).

Whilst the employment status categories relating to whether the individual was an employer were not relevant in this instance as all respondents to the WERS SEQ were employees, managers can receive a different NS-SEC categorisation depending on the number of employees in the workforce. This meant that it was necessary to combine the WERS Management and Employee Surveys to derive an employment status variable for employees before matching on the NS-SEC data. It was also necessary to identify supervisors from the SEQ.

## Female and male CAMSIS scores

CAMSIS scores represent an occupational group's relative position within the national order of social interaction and stratification. As CAMSIS scores are derived separately for men and women and are gender-specific, the variable fcamsis denotes CAMSIS scores for women, whilst mcamsis applies only to men. Further information on CAMSIS scores is available at: <http://www.camsis.stir.ac.uk/index.html>.

## **Missing values**

In 251 cases, xsoc2000 was not coded in the WERS SEQ dataset. As a result, for these employees it was not possible to provide any information on the composition of their occupation, and these cases are labelled as 'occupation unknown' on the derived variables.

The NS-SEC variable contains an additional missing category as it was not possible to assign a value in 283 cases where it was not known whether the employee was a supervisor. As the CAMSIS scores are gender-specific, there are two additional missing codes relating to these derived variables, one to identify cases where the gender of the employee is unknown and the other to identify men on fcamsis, or women on mcamsis.

## **Data sources**

To maximize the number of respondents from any unit group occupation, the four quarterly LFS datasets spanning the period from March 2004 to February 2005 were combined to create an annual database, referred to above as the 2004 LFS. As the LFS re-contacts individual respondents for five consecutive waves, in constructing the annual database, only those who were surveyed for the first or the last time in each quarter were retained so that each individual appears only once.

About one-third of responses to the LFS are given by a proxy on behalf of the intended respondent. The LFS User Guide contains an analysis of the accuracy of proxy responses and concludes that the benefits of obtaining a high response rate (by the use of proxies) outweigh the disadvantages of the errors introduced.

Occupation was defined as the individual's occupation in their main job. Having combined the quarterly LFS datasets and restricted the analysis to employees resident in Great Britain, the smallest occupational group contained 10 individuals. Thirty-seven of the 353 occupations consisted of less than 50 individuals, whilst 86 contained less than 100 employees. The 37 occupations where the occupational derived variables were based on less than 50 cases are flagged on the variable fewLFS so that users can identify occupations where the aggregate data may be least reliable. In addition, the variable nLFS details the total number of individuals within the occupational group.

Before deriving the occupational composition variables, the LFS data was weighted using the person weights to ensure that the data is representative of the population as a whole. The one exception to this was the gross hourly wage rate, where income weights were used. As the LFS data is based on a subsample of respondents to the main survey, the variable fewpay identifies cases where the average gross hourly wage rate for an occupation was based on less than 50 responses (this was the case in 147 occupations).

To supplement this, the variable npay gives the number of cases on which the gross hourly wage rate was based.

NS-SECs and CAMSIS scores were downloaded from the GEODE project website. The CAMSIS scores contained in the downloaded files were originally derived from the 1990 SOC, with SOC 1990 then mapped to SOC 2000. As a result, the CAMSIS scores are approximations to the actual scores for each occupation contained in SOC 2000. The documentation to accompany the file (referenced below) gives further details of this process. Besides the NS-SEC and CAMSIS scores the file containing these data contained a variable recording 4-digit SOC 2000 codes and employment status variables, needed to match the NS-SEC and CAMSIS scores to the WERS 2004 data.

## **Acknowledgements**

This set of derived variables was compiled by the WERS 2004 Information and Advice Service ([www.wers2004.info](http://www.wers2004.info)).

The WIAS team acknowledge the Office for National Statistics Social and Vital Statistics Division and the Northern Ireland Statistics and Research Agency Central Survey Unit as originators of the 2004 LFS data, and the UK Data Archive at the University of Essex as the distributor of the data. The team also acknowledges the GEODE project at the University of Stirling, funded by the Economic and Social Research Council (ESRC) through the National Centre for eSocial Science, for providing the NS-SEC and CAMSIS scores.

The WIAS team acknowledges the Department of Trade and Industry, the ESRC, the Advisory, Conciliation and Arbitration Service and the Policy Studies Institute as the originators of the 2004 Workplace Employment Relations Survey data, and the Data Archive at the University of Essex as the distributor of the data. The National Centre for Social Research was commissioned to conduct the survey fieldwork on behalf of the sponsors.

None of the organisations named above bear any responsibility for the derived variables created.

## **References**

- Department of Trade and Industry. Employment Markets Analysis and Research et al., *Workplace Employee Relations Survey, 2004; Cross-Section Survey, 2004 and Panel Survey, 1998-2004: Wave 2* [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], October 2007. SN: 5294.
- Lambert, Paul (2007) CAMSIS scales for Britain, SOC-2000 [computer file]. University of Stirling: GEODE project. 9 January 2007.
- Office for National Statistics (2005). *The National Statistics Socio-economic Classification User Manual*. Basingstoke: Palgrave Macmillan.
- Office for National Statistics. Social and Vital Statistics Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Quarterly Labour Force Survey, March - May, 2004* [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], November 2005. SN: 4998.

- Office for National Statistics. Social and Vital Statistics Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Quarterly Labour Force Survey, June - August, 2004* [computer file]. *2nd Edition*. Colchester, Essex: UK Data Archive [distributor], November 2005. SN: 5043.
- Office for National Statistics. Social and Vital Statistics Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Quarterly Labour Force Survey, September - November, 2004* [computer file]. *3rd Edition*. Colchester, Essex: UK Data Archive [distributor], March 2006. SN: 5086.
- Office for National Statistics. Social and Vital Statistics Division and Northern Ireland Statistics and Research Agency. Central Survey Unit, *Quarterly Labour Force Survey, December 2004 - February 2005* [computer file]. *3rd Edition*. Colchester, Essex: UK Data Archive [distributor], March 2006. SN: 5162.