

2011 Workplace Employment Relations Study

Note to Accompany the Deposited Data

September 2014

This note introduces the main features of the data of the 2011 Workplace Employment Relations Study (WERS). It covers the files available through the UK Data Service as well as those available through the Secure Data Service and is to be read in conjunction with the other documentation deposited with the data files.

Documentation on the topics below has been added to this note since the last edition (dated July 2014):

- Addition of variables to the general use version of the management data (wers11_mq_general use) (pages 2 and 14);
- Amendments to mqsetup.do to take account of the release of the additional variables (page 14); and
- Lifting of the restriction on the data from the Financial Performance Questionnaire (page 14).

1. Survey components and key design features

The 2011 WERS has the following components:

- Survey of Managers (comprising the Employee Profile Questionnaire and the Management Questionnaire)
- Survey of Worker Representatives
- Survey of Employees
- Financial Performance Questionnaire (for trading sector workplaces only)

The questionnaires for each component are included in the survey documentation.

The sample of the 1990, 1998 and 2004 WERS had a freshly selected cross-section sample and a separate 'panel' element. The 'panel' element consisted of a sample of workplaces that participated in the previous study and were approached to take part in an abridged management interview. The samples were always analysed separately. The key design innovation of the 2011 WERS was to integrate the fresh sample and the 'panel' sample for the purposes of fieldwork and analysis. In the field, workplaces in the 'panel' sample were eligible for all four components of the 2011 WERS. Users therefore have access to management, worker representative and employee data for both waves of the 2004-2011 panel.¹ In addition, weights have been devised which enable the panel sample to be combined with the fresh sample to form a combined sample that is cross-sectionally representative. This more than compensates for the smaller size of the fresh sample in 2011, creating a cross-sectionally representative combined sample that is larger than the cross-section sample that was available in 2004. Further information on the sample design can be

¹ Please note that responses from the same establishment but from a different wave of the 'panel' sample are not necessarily from the same individual, nor can individual respondents be matched across waves.

found in the technical note, "The Design and Administration of the 2011 WERS", included as part of the documentation of the archived data.

2. Overview of the deposited data

Two versions of the 2011 WERS have been deposited at the UK Data Archive. A general use version containing anonymised data made available through the UK Data Service and a restricted access version released through the Secure Data Service. The restricted access version contains the following disclosive or potentially disclosive information:

- postcodes; and
- the Inter-departmental Business Register reference numbers.
- ~~Data from the Financial Performance Questionnaire;~~
- ~~region identifiers;~~
- ~~country in which the workplace is located;~~
- ~~industry classification coded to below the Section level of the *Standard Industrial Classification*; and~~

3. Unique identifiers

Each workplace in the data set has a unique identifier (**serno**) that can be used to link responses from managers, worker representatives and employees at the same establishment. Each worker representative is uniquely identified by the combination of **serno** and **waretyp**. Each employee is uniquely identified by the combination of **persid**.

In respect of the 2004-2011 panel, this data deposit contains data from the 2011 fieldwork process only. However the data of the workplaces in the panel sample contain an unique identifier (**w5serial**) which can be used to link these 2011 observations to the 2004 equivalents held in the 2004 Cross-section Survey of Managers data (file xs04_mq deposited separately, SN5294). Users carrying out this linking should use **w5serial** as the link key but should note that the name for **w5serial** in xs04_mq is **serno**).

4. Variable naming conventions

In general, the variable names of the raw data items contained in the deposited data have two parts. The first part consists of a one or two-character prefix that signifies the section of the relevant questionnaire from which the variable arises. The second part gives some sense of the topic covered by the question. Variables arising from questions that permitted multiple responses have a number at the end to signify the order of response. The names of the variables arising from the fieldwork administration system are descriptive while those derived during data editing or primary analysis are generally pre-fixed with the letters 'e', 'n', 'w_' and 'x'. Variables beginning with 'xcode' or 'pxcode' are overcodes included in the dataset to alert users about particular issues (see the section on overcodes below for details).

Variables in wers11_mq_general use and wers11_mq_restricted_access :

The variables in this file have a one-character prefix which signifies the section of the Management Questionnaire from which the variable arises. The name of a variable

follows the name of the question so **asingle** contains data collected from question ASINGLE from Section A of the questionnaire. Variables arising from multiple response questions are numbered from 1 upwards (or, from 01 if 10 or more responses were permitted), so that **ahowch01** contains the first numeric response given by a particular manager to the question about changes of ownership, and **ahowch14** the fourteenth response. Note, however, that few respondents gave the maximum number of responses to any multiple response question; in most cases they mentioned only one or two items from the code list. In these instances, the remaining variables in the set will be empty (i.e. 'system missing').

*Variables in **wers11_wrq_general use** and **wers11_wrq_restricted_access**:*
Variables arising from the Worker Representative Questionnaire have a two-character prefix. The first character (w) is short-hand for Worker Representative. The second character signifies the section of the questionnaire from which the variable arises. Again, the name of a variable follows the name of the question so **wareptyp** holds data collected from question WAREPTYP from Section A of the Worker Representative Questionnaire. Variables arising from multiple-response questions are labelled in the same way as in **wers11_mq**.

*Variables in **wers11_seq_general use** and **wers11_seq_restricted_access**:*
Variables arising from the Survey of Employees have a two-character prefix. The first character (q) signifies that the variable is from the Survey of Employees. The second character points to the relevant section of the Survey of Employees Questionnaire. For questions that have a number of elements to them and in which the respondent is invited to provide an answer for each element or row (e.g. **A5**), letters (a, b etc.) are appended to the variable name to identify the response to each row (e.g. **qa5a** to **qa5c**). Questions inviting more than one box to be ticked (**E7** and **E12**) yield one dichotomous variable for each of the possible responses (e.g. **qe7mul1** to **qe7mul17**). An additional variable with the same name as the question (**qe7** in this example) indicates the number of boxes ticked by the respondent.

Some questions in the Survey of Employees Questionnaire were not intended to elicit multiple responses but were multi-coded by a number of respondents. For questions that attracted multiple responses from more than 0.5% of respondents, the multiple responses have been made available on the dataset. The naming convention of those variables is outlined in the section on multiple response sets below.

*Variables in **wers11_fpq** and **wers11_fpq_restricted_access**:*
The variable names on these files are largely descriptive.

5. Value labelling conventions

Variables which contain data that are drawn solely from a single question (for example, **afranch**) are labelled using the full text of the relevant question if that question is less than 80 characters long. The text of questions that are over 80 characters long is summarised in the relevant variable labels so users will need to refer to the questionnaire for the full wording of those questions.

6. Missing value coding

The standard approach to the coding of missing values on the 2011 WERS is to use the following codes in all the deposited data files.

- -9 - Not answered / Refused

- -8 - Don't know
- -1 - Item not applicable, i.e. not asked
- -6 - single-coded questions in the SEQ that were inadvertently multi-coded by respondents (e.g. **qd2c**)

In order to facilitate the conversion of the data between formats, the codes above have not been assigned as user-defined missing values in deposited data files.

7. Weighting variables

The surveys in the WERS series are based on stratified samples in which the sampling fractions vary across the strata of the sampling matrix. The achieved sample of workplaces is therefore unrepresentative of the population. In addition, employees from small workplaces were over-represented on the Survey of Employees because questionnaires were distributed to all employees in workplaces with 5-25 employees and to only 25 employees in larger workplaces. Furthermore, variable rates of non-response can cause the achieved sample to depart in additional ways from the population it is intended to represent. Weights equal to $1/(\text{probability of selection \& response})$ were computed for use in the analysis to bring the profiles of the achieved samples of workplaces and employees into line with the profiles of the respective populations, thereby removing known biases introduced by the sample selection and response process.

It is advisable to apply weights when the WERS data are used in regression modelling because the specification of an unweighted model is difficult to achieve in practice. In addition, analysts must also account for the clustering of employee observations within workplaces when using the data from the Survey of Employees in modelling. This may be done through the use of multi-level modelling procedures (also known as random-coefficients models) that take account of the hierarchical nature of the data.

Most software packages, including STATA and SPSS, allow users to apply weights in analysis but some of the procedures, for example *iweights* in STATA and the *WEIGHT BY* in SPSS, will not produce correct standard errors when used on the WERS dataset because of the complex sample design. On the other hand, using the *svy* suite of commands in STATA or the Complex Samples module in SPSS on the WERS dataset will ensure that coefficients and standard errors are both correctly computed.

The weights on the deposited data are listed below.

Combined sample weights:

<i>ESTWTNRC</i>	The standard establishment weight to be applied to the cross-section combined sample of Managers.
<i>EMPWTNRC</i>	To be applied to the cross-section Survey of Managers when one wants to produce analyses reflecting the proportion of employees to whom a particular workplace characteristic pertains.
<i>SEQWTNRC</i>	Employee weight variable to be applied to the Survey of Employees.

<i>FPQWTNRC</i>	Weight to be applied to the data from the Financial Performance Questionnaire.
<i>WRQWTNRC</i>	Weight to be applied to the data for worker representatives (union and non-union) from the Survey of Worker Representatives.
<i>PAIRWTNRC</i>	Weight to be applied to the data for paired data of union and non-union reps from the Survey of Worker Representatives.

Panel sample weights:

<i>PQWTNR</i>	Establishment weight to be applied to the Panel Survey.
<i>PQEMPWTNR</i> <i>TUWTNRP</i>	Employee weight to be applied to the Panel Survey Weight to be applied to the data for union representatives when comparing responses from 2004 to those provided in 2011 within panel workplaces that generated union rep interviews at both time points.
<i>SEQWTNRP</i>	Weight to be applied to the data from the Survey of Employees when comparing employees' responses from 2004 with those provided in 2011 within panel workplaces that generated SEQs at both time points.

All the weights are scaled so that the sum of weighted cases is 100. The advantage of scaling the weights in this way is that the weighted base for any statistic indicates the percentage of the population of establishments (or employees for the Survey of Employees) to which the statistic applies. This can provide a useful reference point during analysis.

Further information on the sample design and the methods used to construct the weights are provided in the technical paper, "The Design and Administration of the 2011 WERS".

8. Multiple-response sets

A number of questions permitted multiple responses (e.g. **CFACTORS** in the Management Questionnaire). These questions are identified with the symbol ^ in the Management and Worker Representative Questionnaires. In the data files, the responses to these questions are stored in successive variables (e.g. **cfactor1**, **cfactor2** and so on). The first variable in the set contains the first response mentioned in the interview, the second variable holds the second response, etc.

The example below takes the multiple response set **CFACTORS** from the Survey of Managers (comprising variables **cfactor1- cfactor9** and **xcfact1- xcfact3**) and computes a (1,0) dummy variable named **cfac_age** to identify those workplaces in which the management respondent chose code 5 'Age'.

To create the dummy variable in SPSS, type:

```
do if missing(cfactor1)=0 .  
+ compute cfac_age=any(5,cfactor1 to cfactor9,xcfact1 to xcfact3) .  
end if .
```

To create the dummy variable in STATA, type:

```
egen cfac_age=eqany(cfactor1-cfactor9 xcfact1-xfact3), values(5)  
recode cfac_age 0=. if cfactor1<1
```

The resulting dummy variable has 244 cases coded 1 and 2436 cases coded 0.

Some multiple-response questions allowed the respondent to give answers other than those included on the pre-specified code frame. In such cases, the 'other' answers provided by the respondent have been coded into additional variables beginning with the letter X (e.g. **xcfact1**, **xcfact2**, **xcfact3**). These additional variables follow the main set on the data file. The original 'other' code remains present in the main set of variables. A respondent using the 'Other, please specify' item on CFACTORS (code 9) will thus have a value of 9 in one of the variables **cfactor1- cfactor9** plus a set of values in **xcfact1- xcfact3** that correspond to the codes assigned to their verbatim answers during the data coding and editing process. The original 'other' code should be ignored during analysis to prevent double-counting.

In the Survey of Employees, questions B5 and E7 were designed to accommodate multiple-responses. In the case of B5, which collected information on the impact of the recession on the respondent, an additional variable (**qb5mul**) was added to the data to indicate the number of responses given by each respondent. For E7, the variables (**qe7mul1 to qe7mul17**) indicate whether the employee has chosen each of the available educational qualifications.

In addition, a small number of people gave multiple responses to a number of questions in the Survey of Employees despite the questionnaire indicating that the respondent should tick only one box. In those cases, the main variable is coded -6. For questions that attracted multiple responses from more than 0.5% of respondents (D2c, D2d and D2e), the multiple responses have been made available on the dataset adjacent to the main variable (for example, qd2cmul1 to qd2cmul5). Also included is a variable which indicates the number of boxes ticked by the respondent (eg qd2cmul).

9. Variables indicating the proportion of employees for whom a practice applied

Questions that asked about the proportion of employees to which a practice applied (for example, COTHJOB) could be answered in two ways. Respondents could cite the percentage using the code frame provided (i.e. 'All (100%)', 'Almost all (80-99%)' etc) or they could give the actual number. The second type of answers were coded into a second variable named using the suffix T (COTHJOB T in this case). Users need to combine the response by recoding the T values back into the percentage code frame.

10. Overcodes

Overcodes are variables for identifying cases where a (major) change has been made to the collected dataset during its editing, or where the research team had reason to be concerned about a particular set of responses. They were included in the dataset to alert users about issues with particular cases so they can consider whether or not to include those cases in their analysis. A list of overcodes is provided in the technical paper on the Design and Administration of the 2011 WERS.

11. Notes about individual variables

Variable	Dataset	
control	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
d1whix1	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
dbwhix	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
dmwhix	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
egrvdisa	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
imatpay	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded “item not applicable” on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main

		stage has meant that the data on this item collected during the pilot cannot be used.
ipatpay	wers11_mq_general use and wers11_mq_restricted_access	A small number of cases were coded "item not applicable" on this variable because they were part of the pilot study. Questionnaire changes between the pilot and the main stage has meant that the data on this item collected during the pilot cannot be used.
lupcode_revised	wers11_mq_restricted_access	The postcodes recorded in lupcode_revised are the postcode of the workplace at the time of the interview not the postcode on the IDBR.
ruluref_panel_2010	wers11_mq_restricted_access	Panel workplaces with valid values in this variable are workplace where the 2003 IDBR reference number was no longer present on the IDBR in 2010 but a tracing exercise carried out as part of the sampling process found another unit with precisely matching characteristics (same name, enterprise, postcode etc). For these workplaces, ruluref_panel_2010 holds the 2010 IDBR local unit reference number of the unit with the precisely matched characteristics. In cases where ruluref_panel_2010 contains a valid IDBR reference number, that number can be used for data linking in preference to the 2003 local unit reference number shown in ruluref_panel_2003.
waunionc	wers11_wrq_general use and wers11_wrq_restricted_access	This variable has replaced w_wauni found in the version of wers11_wrq_general use deposited in February 2013. It contains fully backcoded data on the union of the union representative.
w_wauni	wers11_wrq_general use deposited in February 2013	This variable has been replaced by waunionc found in the version of

		wers11_wrq_general use and wers11_wrq_restricted_access deposited in August 2013.
--	--	---

12. Further information

Further information about 2011 WERS, including the First Findings report and key tables are provided on the GOV.UK website at:

<https://www.gov.uk/government/publications/the-2011-workplace-employment-relations-study-wers>.

2011 WERS additions to survey data

January 2014

The file *WERS 2011 panel weight for 2004 SEQs* was added to the deposited 2011 WERS data set in January 2014. This file was deposited to enable users to compile a dataset containing both the 2004 and 2011 WERS data, and to derive many of the variables used in the primary analysis of the 2011 WERS (see page 12).

The file contains data the compilation program needs to produce the weight to be applied to the 2004 WERS data when a user wishes to compare the characteristics or attitudes of employees between 2004 and 2011 in the population of continuing workplaces (seqwtnrp). The weight addresses any workplace-level non-response or attrition biases arising from the selection of workplaces: (a) into both waves of the panel; and (b) into the subset of these workplaces that returned SEQs in both waves. It also post-stratifies the full set of SEQs in those workplaces by gender.

~~Please note that the the compiling program referred to above is not ready for depositing at the UK Data Archive but will be deposited as soon as possible.~~

2011 WERS additions to survey data

July 2014

The following sets of syntax files and documentation were added to the deposited 2011 WERS data set in July 2014.

Syntax and documentation for compiling a dataset containing both the 2004 and 2011 WERS data and derives many of the variables used in the primary analysis of the 2011 WERS:

- master.do
- label_missing.do
- merging.do
- mqsetup.do
- wrq_setup.do
- seq_set-up.do
- WERS2011 userguide for syntax to set up primary analysis datasets v2 5 March 2014.

Syntax and documentation for replicating the analysis published in the [2011 WERS First Findings](#):

- FF workplaces in the shadow of recession.do;
- FF more work less pay.do;
- FF staffing the workplace.do;
- FF who manages employment relations.do;
- FF what has happened to employee representation.do;
- FF representing employees in the workplace.do;
- FF engaging employees.do;
- FF workplace change.do;
- FF who sets pay and conditions.do;
- FF paying for performance.do;
- FF discontent in the workplace.do;
- FF pay dispersion and satisfaction.do;
- FF long working hours.do;
- FF work-life balance.do;
- FF equality and diversity.do;
- FF trends in training.do;
- FF health and safety.do;
- FF job satisfaction and well-being.do;
- table_prog.do; and
- WERS2011_UserGuide_FFsyntax.

Explanation on the purpose of syntax files can be found in the user guide listed with the files.

All the syntax is in Stata format. There is no equivalent SPSS syntax, as the 2011 WERS Research Team's primary analysis was conducted wholly in Stata. However the Research Team have added a 2011 time-point to the SPSS-based WERS Time Series Dataset (UK Data Service SN: 4511) which is expected to be available soon.

2011 WERS additions to survey data

September 2014

To add further protection to the data when they were first released, regional and country identifiers, detailed industry codes and data from the Financial Performance Questionnaire were released in the restricted access version only. The WERS Steering Committee lifted the restriction on those data items in the summer of 2014 so they were included in the general use version of the data deposited in September 2014.

The release of the detailed industry codes has meant that the bank of syntax for deriving some variables relating to NFSOC (method of pay determination for the largest occupational group) on the syntax file mqsetup.do can now be run on the general use version of the data. A version of mqsetup.doc, updated to take account of the general release of the detailed industry codes, has been included in the material deposited in September 2014. Full details about this can be found in the user guide for the syntax to set up the primary analysis datasets deposited with the syntax files.