

## USER GUIDE FOR 2011 WERS RESEARCH TEAM SYNTAX FILES

### 1. Introduction

This document accompanies the Stata syntax files made available by the 2011 WERS Research Team. This suite of syntax files compiles a dataset containing both the 2004 and 2011 WERS data, and also derives many of the variables used in the primary analysis of the 2011 WERS.

Users should note that, in making this syntax publicly available, we have not sought to prove the veracity of every derived variable; the syntax is the product of the 2011 WERS Research Team's work and has not been independently verified.

All 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 researchers at NIESR have added a 2011 time-point to the SPSS-based WERS Time Series Dataset (UK Data Service SN: 4511).

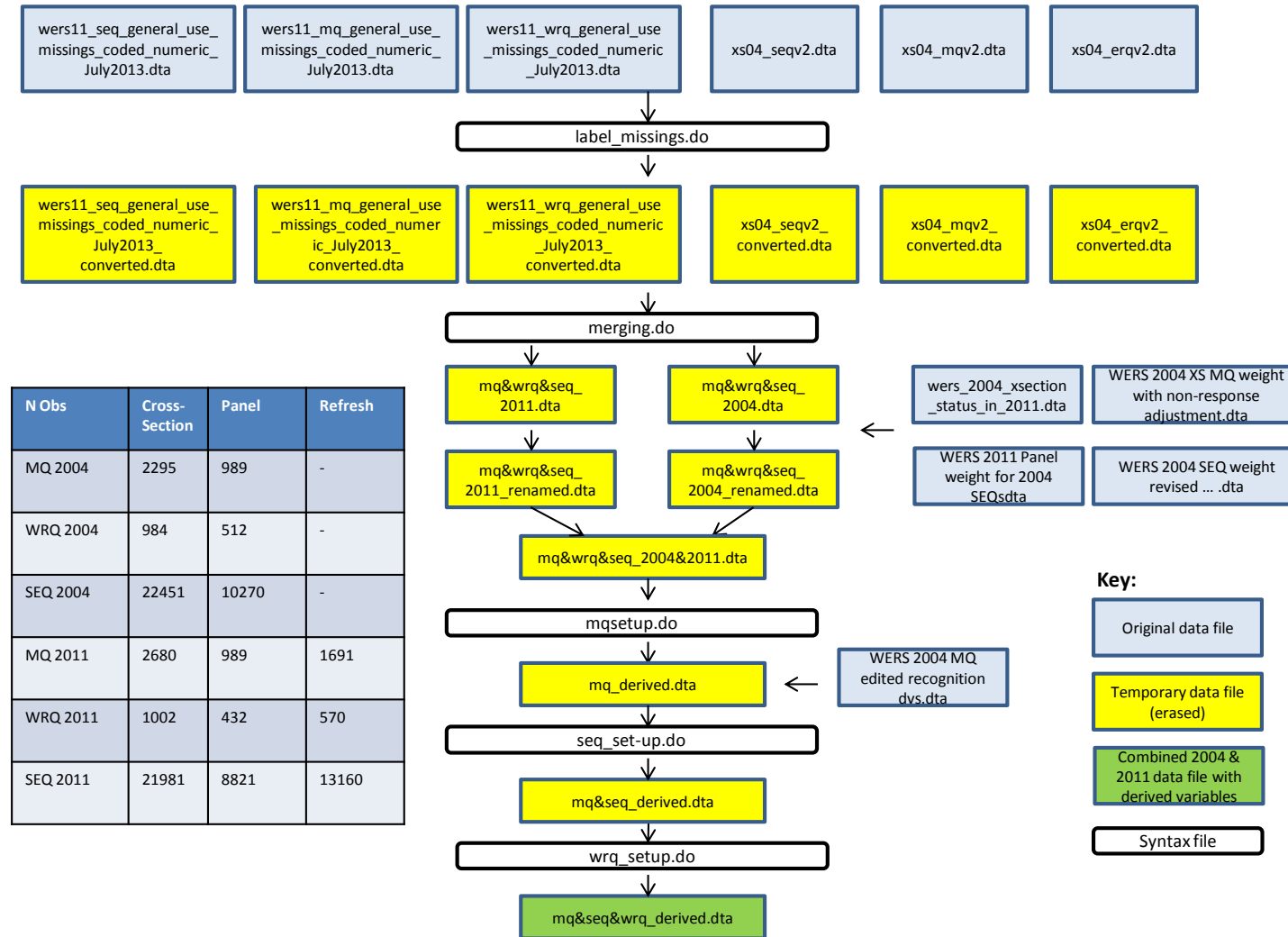
### 2. Overview

The Stata syntax to create the 2011 WERS Research Team's working dataset is contained in a series of separate do files:

- [master.do](#);
- [label\\_missings.do](#);
- [merging.do](#);
- [mqsetup.do](#);
- [seq\\_set-up.do](#); and
- [wrq\\_setup.do](#).

The 'master' do file calls each of the other do files in turn to create a combined 2004 and 2011 dataset and construct the derived variables. A diagram of the production process is provided below in Figure 1. The individual syntax files are described in more detail below.

Figure 1: Data production process



master.do

The master do file calls each of the other syntax files in turn in order to create the final dataset. In order to run the master do file, it is first necessary to change the file paths to those used on your own computer. The master file refers to five different file paths as listed in Table 1.

**Table 1: List of file paths in *master.do***

path_2004_raw	The location where you have stored the 2004 WERS data provided by the UK Data Service
path_2011_raw	The location where you have stored the 2011 WERS data provided by the UK Data Service
path_extradata	In order to run the full set-up syntax, the additional data files listed below provided by the UK Data Service are required and should be saved to this location.  Data files lodged with study number 5294: <ul style="list-style-type: none"> <li>• WERS 2004 XS MQ weight with non-response adjustment.dta</li> <li>• WERS 2004 MQ edited recognition dvs.dta</li> <li>• WERS 2004 XS SEQ weight revised to incorporate new MQ weight with non-response adjustment.dta</li> </ul> Data files lodged with study number 7226: <ul style="list-style-type: none"> <li>• WERS 2011 panel weight for 2004 SEQs</li> </ul>
path_derived	The location where the resulting combined 2004 and 2011 dataset will be saved.
path_log	The location where all resulting log files will be saved
path_do	The location where the do files are stored

Once these file paths have been set correctly in the master file, those using the 'general use' data files provided by the UK Data Service will not need to make any further amendments to any of the do files; they simply need to run 'master.do'.

Those using the restricted data files in the Secure Data Service or the Virtual Microdata Lab will, however, also have to alter the names of the original deposited data files within 'label\_missings.do' as the SDS versions of the deposited data have their own specific file names.

label\_missings.do

Takes each of the 2004 and 2011 data files in turn and recodes all user-defined missing values to Stata format. Labels all missing value codes and saves each data file with the name *\*\_converted.dta*.

### merging.do

This merges the 2004 Management Questionnaire (MQ), Worker Representative Questionnaire (WRQ) and Survey of Employees Questionnaire (SEQ) and the 2011 MQ, WRQ and SEQ. It matches on the 2011 equivalent of the 2004 serno and appends the 2004 data to the 2011 data. It derives variables to identify observations from each of the individual datasets (listed below) and saves a file containing all the 2004 and 2011 data (*mq&wrq&seq\_2004&2011.dta*).

### mqsetup.do

This file derives MQ-based variables from the combined 2004 and 2011 data created in *merging.do*. and saves a version of the combined data file which includes these MQ-derived variables (*mq\_derived.dta*).

The version of mqsetup.do deposited in July 2014 commented out a bank of syntax for deriving some variables relating to NFSOC (method of pay determination for the largest occupational group) because it calls on the detailed 5-digit industry codes which were not released on the general use datafiles. The detailed industry codes have now been released and mqsetup.do has been updated to remove the comments between lines 3317 and 4145.

The WERS Research Team used commands in the Stata module *fre* to display one-way frequency tables in this file to check some of the derived variables. If you do not have the module loaded onto your system, you can download the module using this link <http://ideas.repec.org/c/boc/bocode/s456835.html> or comment out the syntax lines starting with “fre”.

Some of the 2004 WERS variables used in this file are named in capital letters in the data files but referenced in lowercase letters in the do files. This has generated error messages saying that those variable do not exist when mqsetup is run on some platforms. If that happens, change the code to capital letters.

### seq\_set-up.do

This file derives SEQ-based variables from the combined 2004 and 2011 data created in *mqsetup.do*. and saves a version of the combined data file which includes these SEQ-derived variables (*mq&seq\_derived.dta*).

The WERS Research Team used commands in the Stata module *fre* to display one-way frequency tables in this file to check some of the derived variables. If you do not have the module loaded onto your system, you can download the module using this link <http://ideas.repec.org/c/boc/bocode/s456835.html> or comment out the syntax lines starting with “fre”.

### wrq\_setup.do

This file derives WRQ-based variables from the combined 2004 and 2011 data created in *seq\_set-up.do*. and saves a version of the combined data file which includes these WRQ-derived variables (*mq&seq&wrq\_derived.dta*).

The WERS Research Team used commands in the Stata module *fre* to display one-way frequency tables in this file to check some of the derived variables. If you do not have the module loaded onto your system, you can download the module using this link <http://ideas.repec.org/c/boc/bocode/s456835.html> or comment out the syntax lines starting with “fre”.

### 3. Overview of the data structure of *mq&wrq&seq\_2004&2011.dta*

The suite of do files produces a single dataset with the broad structure outlined in Figure 2 below. The notable feature of the dataset is that MQ, WRQ and SEQ data from both 2004 and 2011 is all held within one single data file. Indicator variables (listed in Table 2) are then used to select specific components from within this large dataset for analysis (e.g. to select MQ records only).

**Figure 2: Data structure of *mq&wrq&seq\_2004&2011.dta***

2004 MQ	2004 WRQ	2004 SEQ	serno	zallemps	...	auktotx	...	hwarning	...	nyear	nmq	nmq2004	nmq2011	Nwrq	nseq
	2004 WRQ	2004 SEQ	1	100		.		1		0	1	1	0	1	1
	2004 WRQ	2004 SEQ	1	100		.		1		0	0	0	0	1	1
		2004 SEQ	1	100		.		1		0	0	0	0	0	1
		2004 SEQ	1	100		.		1		0	0	0	0	0	1
		2004 SEQ	2	200		.		1		0	1	1	0	1	0
		2004 SEQ	3	300		.		2		0	1	1	0	0	0
		2004 SEQ	7	100		3		.		1	1	0	1	1	1
		2004 SEQ	8	100		3		.		1	0	0	0	1	1
2011 MQ	2011 WRQ	2011 SEQ	9	100		3		.		1	0	0	0	0	1
		2011 SEQ	10	100		3		.		1	0	0	0	0	1
		2011 SEQ	11	200		4		.		1	1	0	1	1	0
		2011 SEQ	12	300		5		.		1	1	0	1	0	0

**Table 2: Indicator variables that can be used to select individual components within *mq&wrq&seq\_2004&2011.dta***

Subset	Indicator variable
2004 SEQ	nseq2004=1
2011 SEQ	nseq2011=1
2004 MQ	nmq2004=1
2011 MQ	nmq2011=1
2004 WRQ	nwrq2004=1
2011 WRQ	nwrq2011=1
2004 and 2011 MQ	nmq==1
2004 and 2011 SEQ	nseq==1
2004 and 2011 MQ panel	npanel==1
SEQs from panel workplaces generating 1+ SEQ in both 2004 and 2011	nseqpan==1

One further indicator variable 'nyear' identifies all records from WERS 2004 (nyear==0) or WERS 2011 (nyear==1).

Panel workplaces have their 2004 and 2011 data held on separate records, but these records share the same workplace serial number (serno).

### 4. Weights

Table 3 below lists the names of the main weighting variables on the combined dataset. The weighting variables are simply compiled from those weights that have been made available with

the deposited data files. For any 2004 MQ record, 'nestwt' thus simply contains the weighting values from the final MQ weight provided on the deposited 2004 MQ data file, whilst for any 2011 MQ record it simply contains the weighting values from the final MQ weight provided on the deposited 2011 MQ data file.

**Table3 : Names given to the main weighting variables in *mq&wrq&seq\_2004&2011.dta***

Name	Description	Use
nestwt	MQ establishment weight	Analysis of MQ data from 2004 or 2011: workplace-level (showing the % of workplaces with a particular characteristic)
nempwt	MQ employment weight	Analysis of MQ data from 2004 or 2011: employee-level (showing the % of employees in workplaces with a particular characteristic).
eseqwt	SEQ weight	Analysis of SEQ data from 2004 or 2011
w_wt	WRQ weight	Analysis of WRQ data from 2004 or 2011
npqestwt	MQ establishment weight for the panel	Analysis of MQ data from the 2004-2011 panel sample: workplace-level
npqempwt	MQ employment weight for the panel	Analysis of MQ data from the 2004-2011 panel sample: employee-level

## 5. Variable naming conventions

- In the original deposited data files, all variables from the 2011 SEQ have the pre-fix 'q', whereas no prefix was used in 2004. The *merging.do* syntax file addresses this inconsistency by renaming all of those 2004 SEQ variables which are identical to a variable appearing in the 2011 SEQ. Variable 'a1' in the 2004 deposited SEQ datafile is thus renamed to 'qa1' so as to match the name used for the identical question in the 2011 deposited SEQ data file. Variables which appear in the 2004 SEQ and do not have a 2011 equivalent (such as question 'a5') keep their 2004 name and are retained in the merged data file.
- In the original deposited data files, MQ questions which were identical in 2004 and 2011 have already been given the same variable names. Questions which changed in a substantive way between the 2004 and 2011 surveys have been given an X suffix (e.g. 'auktotx'). Questions from the 2004 MQ which have no 2011 equivalent (e.g. 'aphras01') are again retained in the merged data file.

## 6. Running simple descriptive analyses

The Stata output below provides one example of how the combined data file may be used in analysis. In this example we identify the percentage of workplaces in 2004 with any recognised trade union, and the equivalent figure in 2011, and then test whether the difference between these two cross-sectional estimates is statistically significant from zero. The output shows that, in 2004, 22.1% of workplaces had a recognised union whereas, in 2011, the figure was 21.7%; the difference was not statistically significant from zero at conventional levels ( $p=0.80$ ).

```

-----
. clear all
. set more off
.
. *Open dataset
. use "$path_derived\mq&seq&wrq_derived.dta", clear
.
. *Set weight and psu
. svyset serno [pw=nestwt]

    pweight: nestwt
          VCE: linearized
Single unit: missing
  Strata 1: <one>
    SU 1: serno
    FPC 1: <zero>

.
. * Examine change in union recognition between 2004 and 2011 (workplace weighted)
. svy, subpop(nmq) : mean nanyrecb, over(nyear)
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =          1          Number of obs   =    4847
Number of PSUs   =    3898          Population size =   197.57
                                          Subpop. no. obs =    4847
                                          Subpop. size   =   197.57
                                          Design df      =    3897

    _subpop_1: nyear = 2004
    _subpop_2: nyear = 2011

-----
              |              Linearized
              |              Mean   Std. Err.   [95% Conf. Interval]
-----+-----+-----
nanyrecb     |
  _subpop_1   |   .2210907   .0130325   .1955395   .2466419
  _subpop_2   |   .2166411   .0139522   .1892868   .2439954
-----+-----+-----

. local testvar=word(e(varlist),1)

. local spop1=word(e(over_namelist),1)

. local spop2=word(e(over_namelist),2)

. test [`testvar']`spop1'==[`testvar']`spop2'

Adjusted Wald test

( 1) [nanyrecb]_subpop_1 - [nanyrecb]_subpop_2 = 0

      F( 1, 3897) =    0.06
      Prob > F   =    0.8005
-----

```