Millennium Cohort Study

User Guide to Analysing MCS Data Using SPSS

1st Edition

Elizabeth M. Jones and Sosthenes C. Ketende

November 2010

Centre for Longitudinal Studies

Institute of Education, University of London

Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London
WC1H 0AL

website: www.cls.ioe.ac.uk

ISBN 978-1-906929-17-6

The Centre for Longitudinal Studies (CLS) is a department within the Faculty of Policy and Society of the Institute of Education, University of London. The department houses an ESRC Resource Centre devoted to the collection, management and analysis of large-scale longitudinal data. It is the home of three internationally-renowned birth cohort studies: the 1958 National Child Development Study (NCDS), the 1970 British Cohort Study (BCS) and the Millennium Cohort Study (MCS).

The views expressed in this work are those of the authors and do not necessarily reflect the views of the Economic and Social Research Council, or the consortium of government departments which also contribute to the costs of the Millennium Cohort Study. All errors and omissions remain those of the authors.

Contents

Acknowledgements

We would like to acknowledge the helpful comments we received while constructing this report from our CLS colleagues Lisa Calderwood, Jon Johnson, John McDonald, Samantha Parsons, and Peter Shepherd.

1. Introduction

This document describes how to use the Complex Sample settings in SPSS to analyze data from the Millennium Cohort Study (MCS). It describes how to set up Complex Samples Plan files, analyze data using the CS Plans and perform subgroup analyses using CS Plans. These are shown both through menus and syntax.

The document is aimed at those who are already familiar with SPSS, but not with using its Complex Samples module.

You will need a version of SPSS that has the Complex Samples Module. Before you begin, check that your version of SPSS has this. This document was prepared for SPSS 16.

1.1. The Millennium Cohort Study

The Millennium Cohort Study (MCS) is the fourth of Britain's world-renowned national longitudinal birth cohort studies. Each follows a large sample of individuals born over a limited period of time through the course of their lives, charting the effects of events and circumstances in early life on outcomes and achievements later on. They show how histories of health, wealth, education, family and employment are interwoven for individuals and vary between them.

The MCS Study offers large-scale information about the New Century's children, and the families who are bringing them up, for the four countries of the United Kingdom. Its first wave, carried out during 2001-2002, laid the foundations for a major new longitudinal research resource, taking a new "year long" cohort of around 19,000 babies. In England and Wales cohort members were born over the 12 months starting in September 2000, and over 13½ months from late November 2000 in Scotland and Northern Ireland. Information was collected from parents when the babies were aged nine months. The first survey recorded the circumstances of pregnancy and birth, as well as those of the all-important early months of life, and the social and economic background of the family into which the children have been born. The second survey data were collected when the children were age 3, the third when they reached age 5 and the fourth when they were age 7.

The sample of births selected for the first survey of the MCS was clustered geographically, and disproportionately stratified to over-represent areas with high proportions of ethnic minorities in England, residents of areas of high child poverty and residents of the three smaller countries of the UK. The distribution of the MCS families, including those who joined at MCS2, across strata for each country is given in Table 1 below.

Table 1: MCS Cases by stratum and country

	England		Wales		Scotland		N Ireland		UK	
Strata	***N***	***%***	***n***	***%***	***n***	***%***	***n***	***%***	***n***	***%***
Advantaged	4828	39.49	832	30.14	1145	49.04	723	37.69	7528	39.12
Disadvantaged	4806	39.31	1928	69.86	1191	50.96	1200	62.31	9125	47.42
Ethnic	2591	21.19	n/a	n/a	n/a	n/a	n/a	n/a	2591	13.46
Total	**12225**	**100.00**	**2760**	**100.00**	**2336**	**100.00**	**1923**	**100.00**	**19244**	**100.00**

1.2. Weighting and non-response adjustment

Because of the MCS sample design described above, analysis of MCS data should take into account the sample design features such as clustering, stratification and weighting. If you ignore the MCS sampling design, i.e., if you assume simple random sampling and independence of observations, the standard errors will likely be underestimated and the resulting significance tests invalid. Due to attrition and non-response, weights that also adjust for attrition and non-response are preferable to design weights that adjust only for the unequal sample selection probabilities.

One way of adjusting for possible biases generated by systematic unit non-response is to use non-response weights. Unit non-response in MCS1 and non-response from MCS1 to MCS2 was studied by Plewis (2007). The correlates of non-response for MCS1 and MCS2 were studied and used to produce non-response weights that can be used to adjust for non-response. For MCS2, there are three different types of weights to consider: 1) the sample design weights, 2) the non-response weights at wave 1 which when multiplied by the sample weights produce the overall weights at wave 1 (see Table 11.1 in The Millennium Cohort Study: Technical Report on Sampling, 4th Edition, July 2007, Edited by Ian Plewis) and 3) the non-response weights at wave 2 which when multiplied by the overall weights at wave 1 produce the overall weights at wave 2 (see Table 3 in Plewis (2007) for the mean and standard deviation of these weights by stratum for whole UK analyses as well as further technical details on their calculation). Note that the sample at wave 2 was supplemented by ‘new’ families who were eligible at wave 1, but excluded because their addresses held by the Child Benefit Office were not up to date. For these new families, their non-response weight at wave 2 is defined to be one. There were 97 wave 2 productive families that were not used to generate non-response weights due to missing data on the variables used in the response model. These 97 productive families were given a non-response weight of 1.

Weighting methods to compensate for attrition are available for monotone patterns of non-response. For a monotone pattern, a sequential weighting procedure is typically used. The longitudinal weight at wave 1 is defined as the sample (design) weight. For each wave thereafter, the longitudinal weight is the product of the longitudinal weight at the previous wave multiplied by a non-response weight for the current wave. Typically, at each wave the non-response weight is the estimated inverse of the probability of responding based on a logistic regression model. These logistic models use data from previous waves to predict response at the current wave. However, for non-monotone patterns of non-response, some cases have missing data for previous waves and therefore the standard approach cannot be easily applied. For MCS, 1,444 unproductive families at MCS2 were recovered at MCS3, thus yielding a non-monotonic pattern of non-response. There was also non-monotonic non-response at MCS4.

In order to calculate non-response weights for MCS3 and MCS4, multiple imputation was used to impute the required missing data at waves 2 and 3 for the logistic regression model for the probability of responding. With the missing data ‘filled in’, the pattern of non-response was monotone and then the standard sequential weighting procedure could be used to estimate non-response weights. Note that imputation of missing values was only done for variables found in earlier non-response analyses to be related to non-response, not for all variables in the MCS2 with missing values! Multiple imputation was used to impute missing values at wave 2 due to unit non-response for unproductive cases and item non-response for productive cases. For example, for the 1,444 unproductive families at MCS2 which were recovered at MCS3, missing housing tenure at MCS2 was imputed using their housing tenure at MCS1 and MCS3 along with other predictor variables in the imputation model. We expect that the imputation of missing values of housing tenure at MCS2 to be `good’ as the imputation model `loosely speaking’ involves `interpolation’ of the values at MCS1 and

MCS3. Further detail on the non-response predictor variables and imputation models used are provided in the Third Edition of the Technical Report on Response (Ketende, SC. 2010).

At wave 3 all families in the MCS 'active' sample, the 1922 families had a non-response adjusted weight at wave 2 and therefore we didn't have to deal with missing weights at wave 2. As a result of using multiple imputation, all 18526 issued cases were used in the logistic modeling of response at wave 3. Missing values were imputed 10 times and a logistic model of responding at wave 3 was estimated 10 times, once for each imputed dataset. This yields 10 estimated non-response weights at wave 3 and the weights issued for wave 3 are the average of the 10 weights. For further details, see McDonald and Ketende (2010).

The MCS datasets include weights for various types of analysis depending on the wave(s) of data being used and whether the analysis is confined to data relating to a single country or covers all countries of the UK. These weights and the situations in which they may be used are listed in Table 2.

Table 2: Attrition /non-response weights

MCS data to be analysed	All of UK/ Country-Specific[1] Analyses	Weight variable name[2]
MCS1	Country-specific	aovwt1
MCS1	All of UK	aovwt2
MCS2 or MCS2 in combination with previous wave	Country-specific	bovwt1
MCS2 or MCS2 in combination with previous wave	All of UK	bovwt2
MCS2 or MCS2 in combination with previous wave	GB only	bovwtgb
MCS3 or MCS3 in combination with any previous wave(s)	Country-specific	covwt1
MCS3 or MCS3 in combination with any previous wave(s)	All of UK	covwt2
MCS3 or MCS3 in combination with any previous wave(s)	GB only	covwtgb
MCS4 or MCS4 in combination with any previous wave(s)	Country-specific	dovwt1
MCS4 or MCS4 in combination with any previous wave(s)	All of UK	dovwt2
MCS4 or MCS4 in combination with any previous wave(s)	GB only	dovwtgb

Notes:
[1]Country-specific analyses include analyses for a single country, e.g. Wales and bivariate analyses in which one of the variables is country.
[2]The weights are longitudinal, i.e. they include sampling weight and attrition/ non-response adjustment for all the previous waves.

There are two other weighting variables not listed in the table above namely **weight1** and **weight2** which are the design weights. These can be used in cases where you don't want to adjust for attrition/ non-response.

1.3. MCS data analysis using SPSS

As mentioned previously, the data are stratified and clustered by the type of electoral ward individuals lived in at the time of sampling and this must be taken into account when analysing data. There are a number of ways to take the sample design into account:
1) use SPSS Complex Samples procedure,
2) use analysis procedures that allow robust cluster options for the calculation of robust standard errors, or
3) use a multi-level (hierarchical) model.

The preferable method in SPSS is to use the Complex Samples procedure, which requires the following sample design variables:

Stratification variable. MCS is stratified by design. There are 9 different strata with all UK countries having two strata i.e. advantaged and disadvantaged. England has one more strata for Ethnic minorities. The stratum variable is called **pttype2**.

Clustering variable. MCS is also clustered at electoral ward level. Wards were the primary sampling unit. A few small wards were combined with a neighbouring ward making what is often referred to as super-wards. The ward variable is called **sptn00**.

Finite Population Correction factor (fpc) variable. When the size of the sample becomes a large fraction of the size of the population we use something called a finite population correction factor (fpc). The finite population correction factor measures how much extra precision we achieve when the sample size become close to the population size. This variable is called **nh2** or **Nh2** in the MCS datasets.

1.4. Subgroup analyses

You may at times want to perform analyses for only a subgroup of the MCS cases rather than the entire data set. When you wish to analyze such a subgroup, for example to run analyses for boys only, it is not advisable to exclude the other cases (e.g., girls) from the analyses by dropping cases or using the 'select' or 'if' commands. Instead you should use a subgroup function in your analyses. The subpopulation function will produce results for only your subgroup of interest without completely dropping the other cases from the analyses. Dropping cases may lead to incorrect estimation of standard errors and p-values.

SPSS allows you to specify a subgroup in your analyses. Instructions on how to do this are below in the syntax and menu sections.

2. Setting up complex samples plan files

Before you can analyze data from a complex sample using SPSS, you need to set up Complex Samples Plan (CSPLAN) file(s). These files contain information about sampling levels, weights, and strata and are accessed by SPSS when you perform complex samples analyses.

When you run complex samples analyses, you tell SPSS which CS Plan file you wish to use for each analysis. The only information that may change during your analyses is the weight that you wish to use. This is because all other survey design features such as the strata and primary sampling unit (PSU) are the same in all kinds of analysis. We suggest setting up and saving one CS Plan file for each MCS weight you will use in your analyses.

3. Setting up CS Plan files with syntax

Below is the basic syntax format to create a CS Plan file. The variables that are used in the creation of your plan are a weight variable (in this example, **weight1**), **pttype2**, **sptn00**, **nh2** and **mcsid**. All but one of these variables will remain the same in all of your CS Plan files for MCS data; only the weight variable will change. You will also need to change the plan file name and the design stage label to reflect the weight being used.

```
CSPLAN ANALYSIS
  /PLAN FILE='csplanfilename.csaplan'
  /PLANVARS ANALYSISWEIGHT=weight 1
  /PRINT PLAN
  /DESIGN STAGELABEL='weight1' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

Sections 3.1 through 3.11 below give the syntax to create CS Plan file for each of the weights shown in Table 2.

3.1. Create CS Plan file for complex sample analysis using aovwt1

```
CSPLAN ANALYSIS
  /PLAN FILE='cspaovwt1.csaplan'
  /PLANVARS ANALYSISWEIGHT=aovwt1
  /PRINT PLAN
  /DESIGN STAGELABEL='aovwt1' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.2. Create CS Plan file for complex sample analysis using aovwt2

```
CSPLAN ANALYSIS
  /PLAN FILE='cspaovwt2.csaplan'
  /PLANVARS ANALYSISWEIGHT=aovwt2
  /PRINT PLAN
  /DESIGN STAGELABEL='aovwt2' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.3. Create CS Plan file for complex sample analysis using bovwt1

```
CSPLAN ANALYSIS
  /PLAN FILE='cspbovwt1.csaplan'
  /PLANVARS ANALYSISWEIGHT=bovwt1
  /PRINT PLAN
  /DESIGN STAGELABEL='bovwt1' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.4. Create CS Plan file for complex sample analysis using bovwt2

```
CSPLAN ANALYSIS
  /PLAN FILE='cspbovwt2.csaplan'
  /PLANVARS ANALYSISWEIGHT=bovwt2
  /PRINT PLAN
  /DESIGN STAGELABEL='bovwt2' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.5. Create CS Plan file for complex sample analysis using bovwtgb

```
CSPLAN ANALYSIS
  /PLAN FILE='cspbovwtgb.csaplan'
  /PLANVARS ANALYSISWEIGHT=bovwtgb
  /PRINT PLAN
  /DESIGN STAGELABEL='bovwtgb' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.6. Create CS Plan file for complex sample analysis using covwt1

```
CSPLAN ANALYSIS
  /PLAN FILE='cspcovwt1.csaplan'
  /PLANVARS ANALYSISWEIGHT=covwt1
  /PRINT PLAN
  /DESIGN STAGELABEL='covwt1' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.7. Create CS Plan file for complex sample analysis using covwt2

```
CSPLAN ANALYSIS
  /PLAN FILE='cspcovwt2.csaplan'
  /PLANVARS ANALYSISWEIGHT=covwt2
  /PRINT PLAN
  /DESIGN STAGELABEL='covwt2' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.8. Create CS Plan file for complex sample analysis using covwtgb

```
CSPLAN ANALYSIS
  /PLAN FILE='cspcovwtgb.csaplan'
  /PLANVARS ANALYSISWEIGHT=covwtgb
  /PRINT PLAN
  /DESIGN STAGELABEL='covwtgb' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.9. Create CS Plan file for complex sample analysis using dovwt1

```
CSPLAN ANALYSIS
  /PLAN FILE='cspdovwt1.csaplan'
  /PLANVARS ANALYSISWEIGHT=dovwt1
  /PRINT PLAN
  /DESIGN STAGELABEL='dovwt1' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.10. Create CS Plan file for complex sample analysis using dovwt2

```
CSPLAN ANALYSIS
  /PLAN FILE='cspdovwt2.csaplan'
  /PLANVARS ANALYSISWEIGHT=dovwt2
  /PRINT PLAN
  /DESIGN STAGELABEL='dovwt2' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

3.11. Create CS Plan file for complex sample analysis using dovwtgb

```
CSPLAN ANALYSIS
  /PLAN FILE='cspdovwtgb.csaplan'
  /PLANVARS ANALYSISWEIGHT=dovwtgb
  /PRINT PLAN
  /DESIGN STAGELABEL='dovwtgb' STRATA=pttype2 CLUSTER=sptn00
  /ESTIMATOR TYPE=EQUAL_WOR
  /POPSIZE VARIABLE=nh2
  /DESIGN CLUSTER=mcsid
  /ESTIMATOR TYPE=EQUAL_WOR
  /INCLPROB VALUE=1.0.
```

4. Analyzing data using CS Plan files with syntax

Once you have set up and saved your CS Plan file(s), they are ready to be used in your analyses, but make sure that you select the correct CS Plan file for the type of analysis you are doing. You will need to specify a CS Plan for each command you run. This can be done by putting the letters 'CS' at the start of your simple random sample equivalent command, and putting a /PLAN FILE subcommand identifying the CS Plan file you wish to use, as in the example below.

```
CSLOGISTIC outcome (LOW) BY factor1 factor2 … factorn WITH covar1 covar2 …
covarn
  /PLAN FILE='csplanfilename.csaplan'
  /MODEL BY factor1 factor2 … factorn covar1 covar2 … covarn
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE]
LCONVERGE=[0] CHKSEP=20 CILEVEL=95
  /PRINT SAMPLEINFO.
```

5. Performing subgroup analyses with complex samples with syntax

To use a subgroup, simply add a /DOMAIN VARIABLE subcommand that identifies your subgroup with the category to be included in parentheses, as shown below.

If you wish to select a subpopulation based on multiple variables, for example to run analyses for boys whose mothers are employed, you should first create a dummy variable that identifies your group of interest and use that dummy in the subpopulation option.

```
CSLOGISTIC outcome (LOW) BY factor1 factor2 … factorn WITH covar1 covar2 …
covarn
  /PLAN FILE='csplanfilename.csaplan'
  /DOMAIN VARIABLE= dummy(1)
  /MODEL BY factor1 factor2 … factorn covar1 covar2 … covarn
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER SE CINTERVAL TTEST
  /TEST TYPE=F PADJUST=LSD
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE]
LCONVERGE=[0] CHKSEP=20 CILEVEL=95
  /PRINT SAMPLEINFO.
```

6. Examples of analyses using complex samples

Below are a series of examples of using Complex Samples for running analyses of MCS data.

The lines below simply merge together two files to create the active data file for analysis. The files are the longitudinal family file and the MCS 4 parent interview file.

```
GET FILE="mcs4_parent_interview.sav".
SORT CASES BY mcsid(A).
SAVE OUTFILE="mcs4_parent_interview.sav".

GET FILE="mcs_longitudinal_family_file.sav".
SORT CASES BY mcsid(A).
SAVE OUTFILE="mcs_longitudinal_family_file.sav".

GET FILE="mcs4_parent_interview.sav".

MATCH FILES /TABLE=*
  /FILE="mcs_longitudinal_family_file.sav"
  /RENAME (dovwt1 dovwt2 weight1 weight2 daoutc00 = d0 d1 d2 d3 d4)
  /BY  mcsid
  /DROP= d0 d1 d2 d3 d4.
```

Here we select only those families who were productive at MCS 4.

```
SELECT IF daoutc00 EQ 1.
```

6.1. Means

The following lines produce means for the SDQ Total Difficulties scale over child gender. The “CS” as a prefix to the DESCRIPTIVES command indicates that Complex Samples analysis should be used. The PLAN FILE subcommand indicates which CS Plan file should be used. Here we have used the plan file that contains the weight dovwt2.

```
* Means of SDQ total difficulties over child gender.
CSDESCRIPTIVES
  /PLAN FILE='csps4dovwt2.csaplan'
  /SUMMARY VARIABLES=ddebdta0
  /SUBPOP TABLE=dhcsexa0 DISPLAY=LAYERED
  /MEAN
  /STATISTICS COUNT POPSIZE
  /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.
```

Here are the results of the above command. You can see that both the weighted population size and unweighted counts are given.

Univariate Statistics

		Estimate	Population Size	Unweighted Count
Mean	S4 DV SDQ Total Difficulties C1	7.68	13410.902	13363

Subpopulation Descriptives

Univariate Statistics

S4 HHQ CM1 Cohort Member Sex			Estimate	Population Size	Unweighted Count
Male	Mean	S4 DV SDQ Total Difficulties C1	8.36	6898.304	6793
Female	Mean	S4 DV SDQ Total Difficulties C1	6.96	6512.598	6570

6.2. Cross-tabulations

The lines below run a variety of crosstabulations. Again, the “CS” prefix is used with the TABULATE command, and the PLAN FILE subcommand is used. The output generated follows each command.

```
* Tabulation of exercise by child gender.
CSTABULATE
  /PLAN FILE='csps4dovwt2.csaplan'
  /TABLES VARIABLES=dmplfra0 BY dhcsexa0
  /CELLS COLPCT
  /STATISTICS COUNT
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

S4 MAIN Frequency physical activity C1 * S4 HHQ CM1 Cohort Member Sex

			S4 HHQ CM1 Cohort Member Sex		
S4 MAIN Frequency physical activity C1			Male	Female	Total
Five or more days a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	67.3%	64.2%	65.8%
		Unweighted Count	4681	4311	8992

Four days a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	5.6%	6.0%	5.8%
		Unweighted Count	394	394	788
Three days a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	7.2%	8.2%	7.7%
		Unweighted Count	515	565	1080
Two days a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	7.7%	8.5%	8.1%
		Unweighted Count	557	569	1126
One day a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	5.4%	6.3%	5.8%
		Unweighted Count	385	456	841
Less often than once a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	1.8%	2.0%	1.9%
		Unweighted Count	121	135	256
Not at all	% within S4 HHQ CM1 Cohort Member Sex	Estimate	4.9%	4.9%	4.9%
		Unweighted Count	346	355	701
Total	% within S4 HHQ CM1 Cohort Member Sex	Estimate	100.0%	100.0%	100.0%
		Unweighted Count	6999	6785	13784

```
* Tabulation of games console access by child gender.
CSTABULATE
  /PLAN FILE='csps4dovwt2.csaplan'
  /TABLES VARIABLES=dmcmgaa0 BY dhcsexa0
  /CELLS COLPCT
  /STATISTICS COUNT
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

S4 MAIN Does CM have access to games system (ex handheld consoles) C1 * S4 HHQ CM1 Cohort Member Sex

S4 MAIN Does CM have access to games system (ex handheld consoles) C1			S4 HHQ CM1 Cohort Member Sex: Male	Female	Total
Yes	% within S4 HHQ CM1 Cohort Member Sex	Estimate	79.4%	58.6%	69.3%
		Unweighted Count	5587	3915	9502
No	% within S4 HHQ CM1 Cohort Member Sex	Estimate	20.6%	41.4%	30.7%
		Unweighted Count	1411	2869	4280
Total	% within S4 HHQ CM1 Cohort Member Sex	Estimate	100.0%	100.0%	100.0%
		Unweighted Count	6998	6784	13782

```
* Tabulation of chores by child gender.
CSTABULATE
  /PLAN FILE='csps4dovwt2.csaplan'
  /TABLES VARIABLES=dmachma0 BY dhcsexa0
  /CELLS COLPCT
  /STATISTICS COUNT
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

S4 MAIN How often is CM involved in household chores C1 * S4 HHQ CM1 Cohort Member Sex

S4 MAIN How often is CM involved in household chores C1			S4 HHQ CM1 Cohort Member Sex: Male	Female	Total
Every day or almost every day	% within S4 HHQ CM1 Cohort Member Sex	Estimate	29.2%	33.7%	31.4%
		Unweighted Count	2023	2237	4260

Several times a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	18.7%	21.8%	20.2%
		Unweighted Count	1288	1498	2786
Once or twice a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	27.6%	26.8%	27.2%
		Unweighted Count	1908	1855	3763
Once or twice a month	% within S4 HHQ CM1 Cohort Member Sex	Estimate	8.8%	7.4%	8.1%
		Unweighted Count	629	488	1117
Less often than once a month	% within S4 HHQ CM1 Cohort Member Sex	Estimate	5.1%	4.0%	4.6%
		Unweighted Count	361	263	624
Not at all	% within S4 HHQ CM1 Cohort Member Sex	Estimate	10.6%	6.4%	8.6%
		Unweighted Count	787	442	1229
Total	% within S4 HHQ CM1 Cohort Member Sex	Estimate	100.0%	100.0%	100.0%
		Unweighted Count	6996	6783	13779

6.3. Logistic regression

Below is an example of a logistic regression using Complex Samples. Again, the prefix ‘CS’ is put before the LOGISTIC command. This example uses games console access (yes/no) as an outcome, and cohort member age, mother’s education, and country of residence at interview as predictors.

```
CSLOGISTIC  dmcmgaa0(HIGH) BY  dmdnvq00 ddcnty00 WITH dhcagea0
  /PLAN FILE='csps4dovwt2.csaplan'
  /MODEL dhcagea0 dmdnvq00 ddcnty00
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER TTEST
  /TEST TYPE=F PADJUST=LSD
  /ODDSRATIOS FACTOR=[dmdnvq00(96)]
  /MISSING CLASSMISSING=EXCLUDE
```

```
  /CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE]
LCONVERGE=[0] CHKSEP=20 CILEVEL=95
  /PRINT NONE.
```

Tests of Model Effects

Source	df1	df2	Wald F	Sig.
(Corrected Model)	10.000	380.000	12.271	.000
(Intercept)	1.000	389.000	.940	.333
dhcagea0	1.000	389.000	5.303	.022
dmdnvq00	6.000	384.000	17.117	.000
ddcnty00	3.000	387.000	6.292	.000

Dependent Variable: S4 MAIN Does CM have access to games system (ex handheld consoles) C1 (reference category = Yes)
Model: (Intercept), dhcagea0, dmdnvq00, ddcnty00

Parameter Estimates

dmcm gaa0	Parameter	B	Hypothesis Test t	df	Sig.
0	(Intercept)	.646	.986	389.000	.325
	dhcagea0	.000	-2.303	389.000	.022
	[dmdnvq00=1]	-.190	-1.598	389.000	.111
	[dmdnvq00=2]	-.290	-3.427	389.000	.001
	[dmdnvq00=3]	-.179	-1.906	389.000	.057
	[dmdnvq00=4]	.228	2.500	389.000	.013
	[dmdnvq00=5]	.515	4.620	389.000	.000
	[dmdnvq00=95]	.127	.786	389.000	.432
	[dmdnvq00=96]	.000[a]	.	.	.
	[ddcnty00=1]	.138	1.717	389.000	.087
	[ddcnty00=2]	-.060	-.649	389.000	.516
	[ddcnty00=3]	-.188	-1.846	389.000	.066
	[ddcnty00=4]	.000[a]	.	.	.

Dependent Variable: S4 MAIN Does CM have access to games system (ex handheld consoles) C1 (reference category = Yes)

Model: (Intercept), dhcagea0, dmdnvq00, ddcnty00

a. Set to zero because this parameter is redundant.

6.4. Subgroup analysis

You can run subgroup analyses for some commands (but not all). Below is an example of our logistic regression, run for families in London only. As in the logistic regression example above, games console access (yes/no) is the outcome, and cohort member age and mother's education are predictors. Country of residence is no longer included, as it will be England for all included cases. The DOMAIN VARIABLE subcommand specifies the subgroup, which has value 7 on the variable ddregn00. Note that dovwt1 is used as the weight, as using the London subgroup limits the cases to those in a single country.

```
CSLOGISTIC  dmcmgaa0(HIGH) BY   dmdnvq00 WITH dhcagea0
  /PLAN FILE='csps4dovwt1.csaplan'
  /DOMAIN VARIABLE=ddregn00(7)
  /MODEL dhcagea0 dmdnvq00
  /INTERCEPT INCLUDE=YES SHOW=YES
  /STATISTICS PARAMETER TTEST
  /TEST TYPE=F PADJUST=LSD
  /ODDSRATIOS FACTOR=[dmdnvq00(96)]
  /MISSING CLASSMISSING=EXCLUDE
  /CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1e-006 RELATIVE]
LCONVERGE=[0] CHKSEP=20 CILEVEL=95
  /PRINT NONE.
```

Parameter Estimates

dmcm gaa0	Parameter	B	Hypothesis Test t	df	Sig.
0	(Intercept)	3.852	1.650	266.000	.100
	dhcagea0	-.002	-1.760	266.000	.080
	[dmdnvq00=1]	-.537	-1.763	266.000	.079
	[dmdnvq00=2]	-.799	-3.676	266.000	.000
	[dmdnvq00=3]	-.788	-2.851	266.000	.005
	[dmdnvq00=4]	.024	.098	266.000	.922
	[dmdnvq00=5]	.000	.002	266.000	.999
	[dmdnvq00=95]	.133	.399	266.000	.690
	[dmdnvq00=96]	.000[a]	.	.	.

Subpopulation: S4 DV ADMIN Interview GOR = London
Dependent Variable: S4 MAIN Does CM have access to games system (ex handheld consoles) C1 (reference category = Yes)
Model: (Intercept), dhcagea0, dmdnvq00

a. Set to zero because this parameter is redundant.

6.5. Country-specific

Below is an example of running analyses for one country only. Unlike the above subgroup examples, it is acceptable to use SELECT to choose one country, as the country-specific weight will adjust for this. Here we temporarily drop all families except those in Scotland and run a cross-tabulation. Note that the PLAN FILE subcommand specifies the plan file that contains the weight dovwt1.

```
* Tabulation of chores by child gender for Scotland only.
TEMPORARY.
SELECT IF ddcnty00 EQ 3.
CSTABULATE
  /PLAN FILE='csps4dovwt1.csaplan'
  /TABLES VARIABLES=dmachma0 BY dhcsexa0
  /CELLS COLPCT
  /STATISTICS COUNT
  /TEST INDEPENDENCE
  /MISSING SCOPE=TABLE CLASSMISSING=EXCLUDE.
```

S4 MAIN How often is CM involved in household chores C1 * S4 HHQ CM1 Cohort Member Sex

			S4 HHQ CM1 Cohort Member Sex		
S4 MAIN How often is CM involved in household chores C1			Male	Female	Total
Every day or almost every day	% within S4 HHQ CM1 Cohort Member Sex	Estimate	28.6%	32.8%	30.7%
		Unweighted Count	226	251	477
Several times a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	17.6%	23.1%	20.3%
		Unweighted Count	145	189	334

Once or twice a week	% within S4 HHQ CM1 Cohort Member Sex	Estimate	30.5%	24.7%	27.6%
		Unweighted Count	247	200	447
Once or twice a month	% within S4 HHQ CM1 Cohort Member Sex	Estimate	8.8%	7.1%	8.0%
		Unweighted Count	82	60	142
Less often than once a month	% within S4 HHQ CM1 Cohort Member Sex	Estimate	4.7%	5.7%	5.2%
		Unweighted Count	40	41	81
Not at all	% within S4 HHQ CM1 Cohort Member Sex	Estimate	9.9%	6.6%	8.3%
		Unweighted Count	78	49	127
Total	% within S4 HHQ CM1 Cohort Member Sex	Estimate	100.0%	100.0%	100.0%
		Unweighted Count	818	790	1608

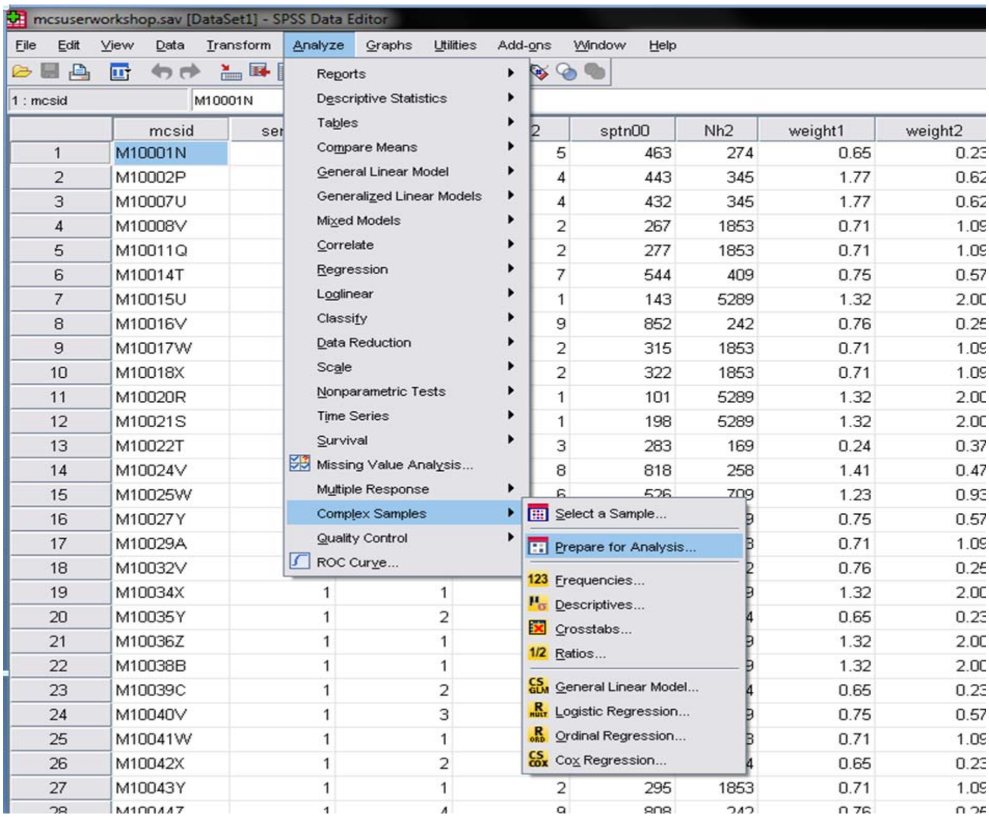
Tests of Independence

		Chi-Square	Adjusted F	df1	df2	Sig.
S4 MAIN How often is CM involved in household chores C1 * S4 HHQ CM1 Cohort Member Sex	Pearson	20.870	3.514	4.448	395.892	.006
	Likelihood Ratio	20.947	3.527	4.448	395.892	.006

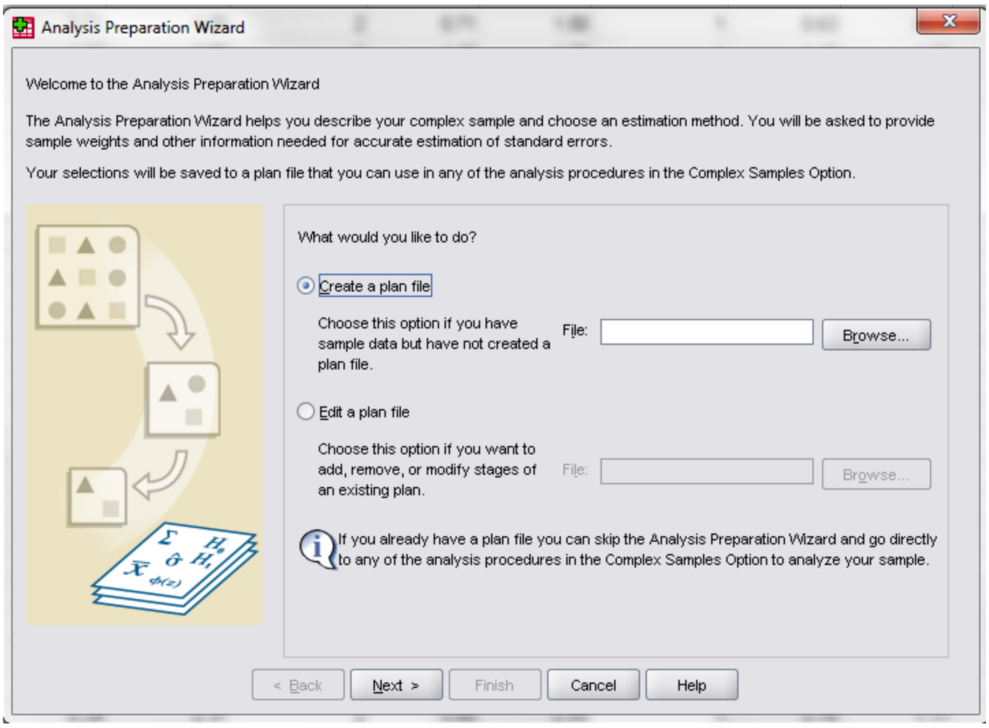
The adjusted F is a variant of the second-order Rao-Scott adjusted chi-square statistic. Significance is based on the adjusted F and its degrees of freedom.

7. Setting up CS Plan files through menus

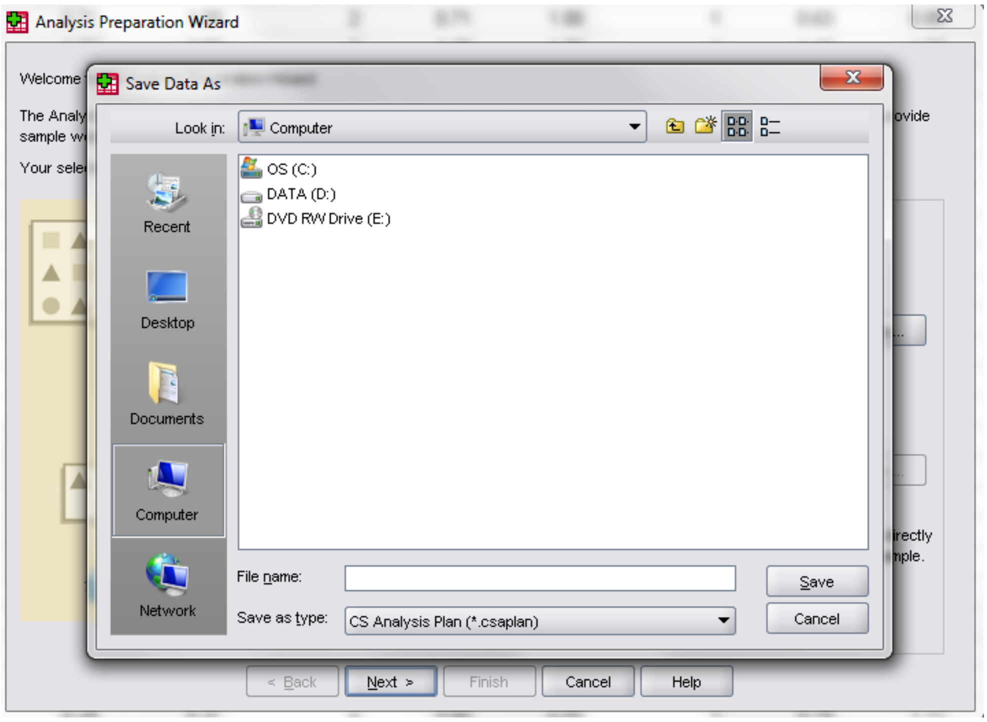
Go to Analyze | Complex Samples | Prepare for Analysis…



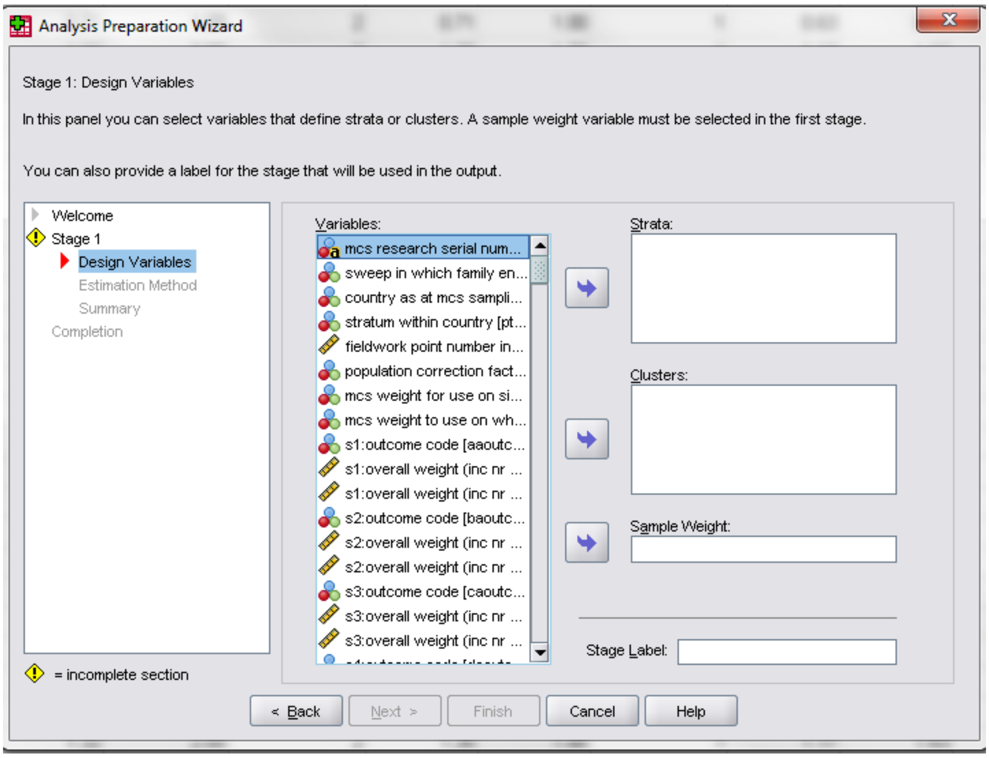
This will open a wizard. Select ***Create a plan file*** and click ***Next***.



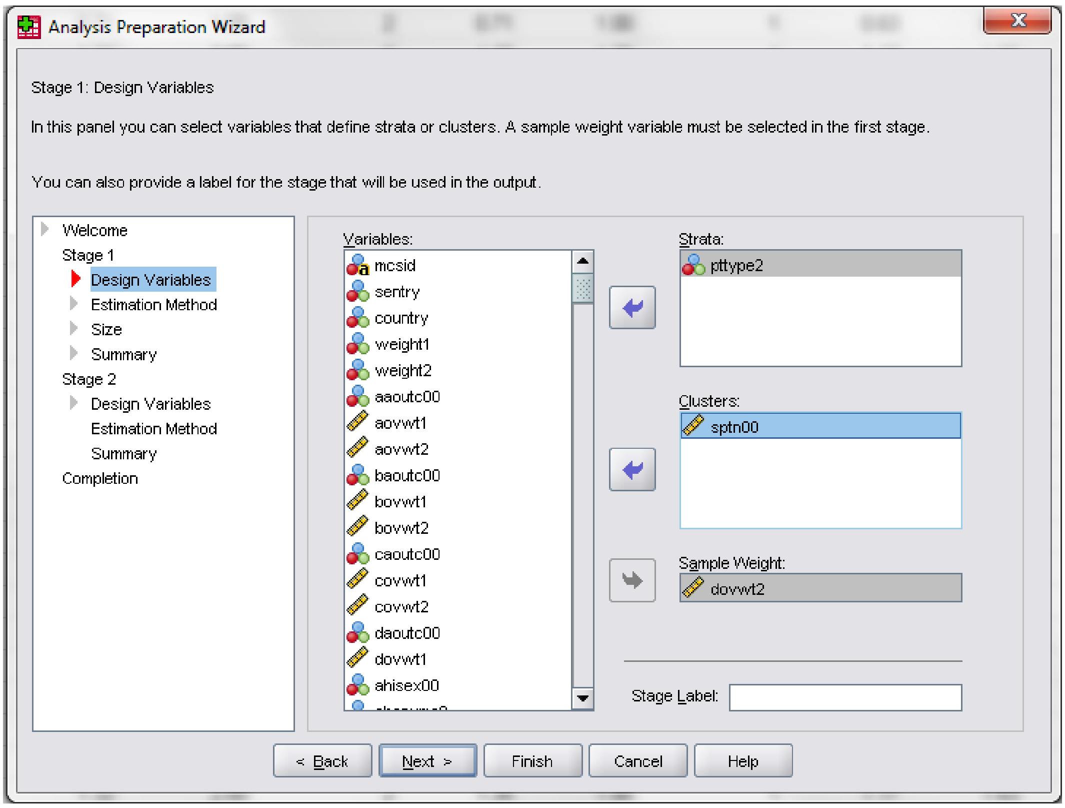
You will then be asked for a location to save the CS Plan file and a name of your CS Plan file. We suggest that you include the name of the weight in the file name as this is the only information that will change from file to file and this will allow you to identify which file to use.



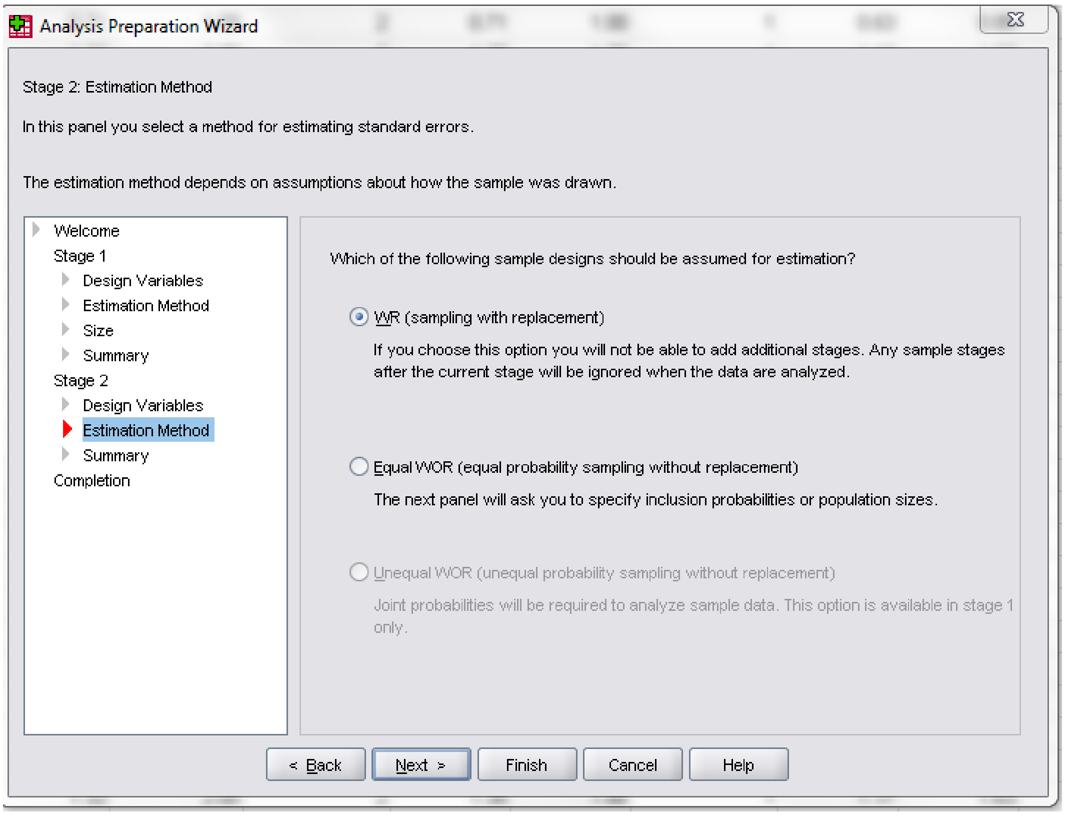
Once you have provided a name, you will be asked for the design variables for the first level of the design.



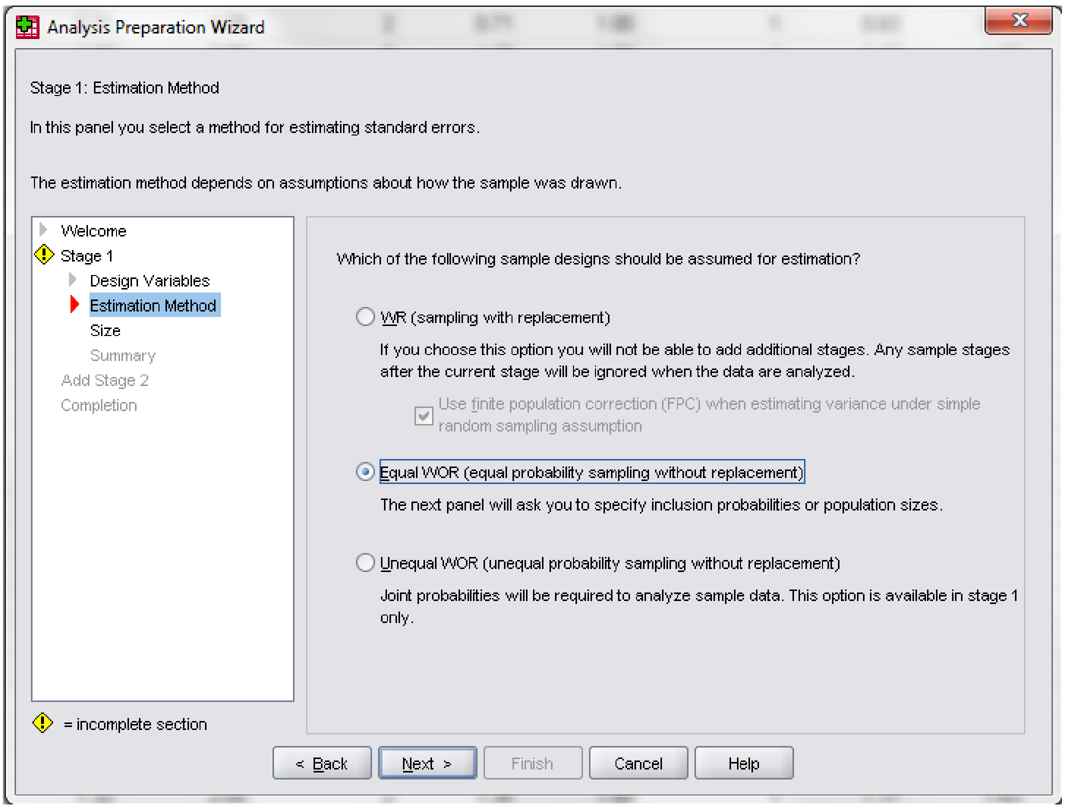
Fill in the information as follows. The ***Strata*** variable is **pttype2** and the ***Clusters*** variable is **sptn00**. You will select the weight you want to use for this CS Plan file, e.g., **dovwt2**. Then click ***Next***.



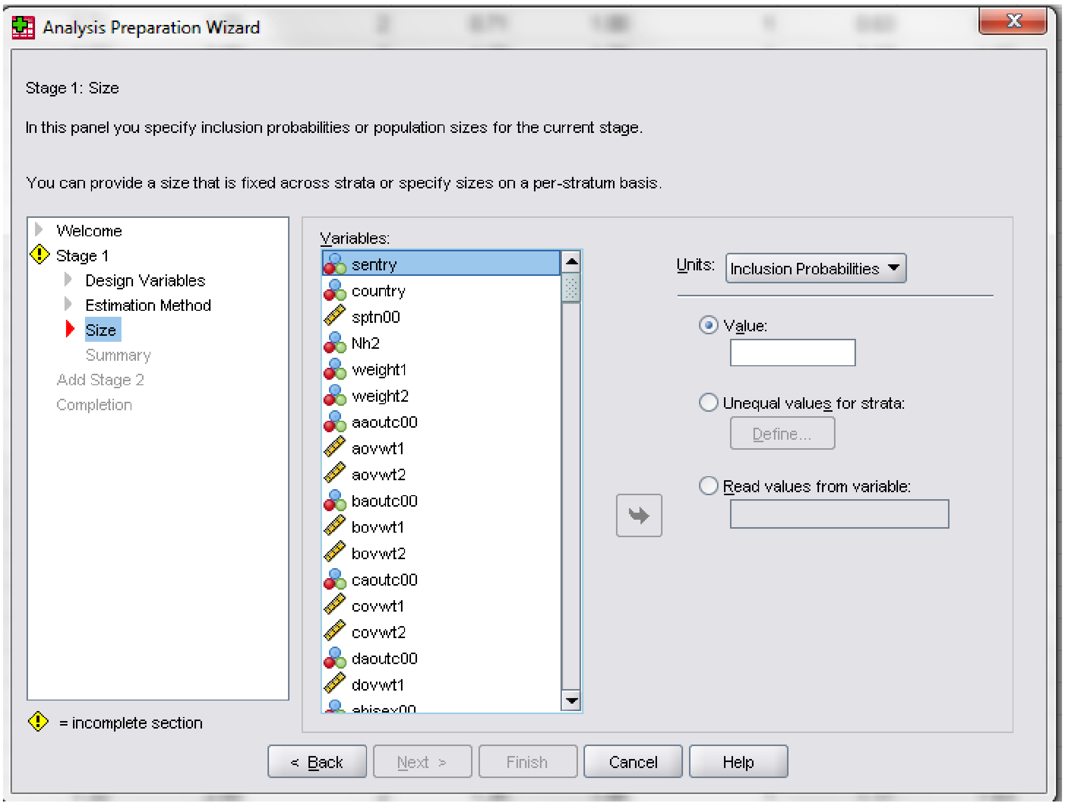
In the next box, ***WR*** will automatically be selected, but you want to change this…



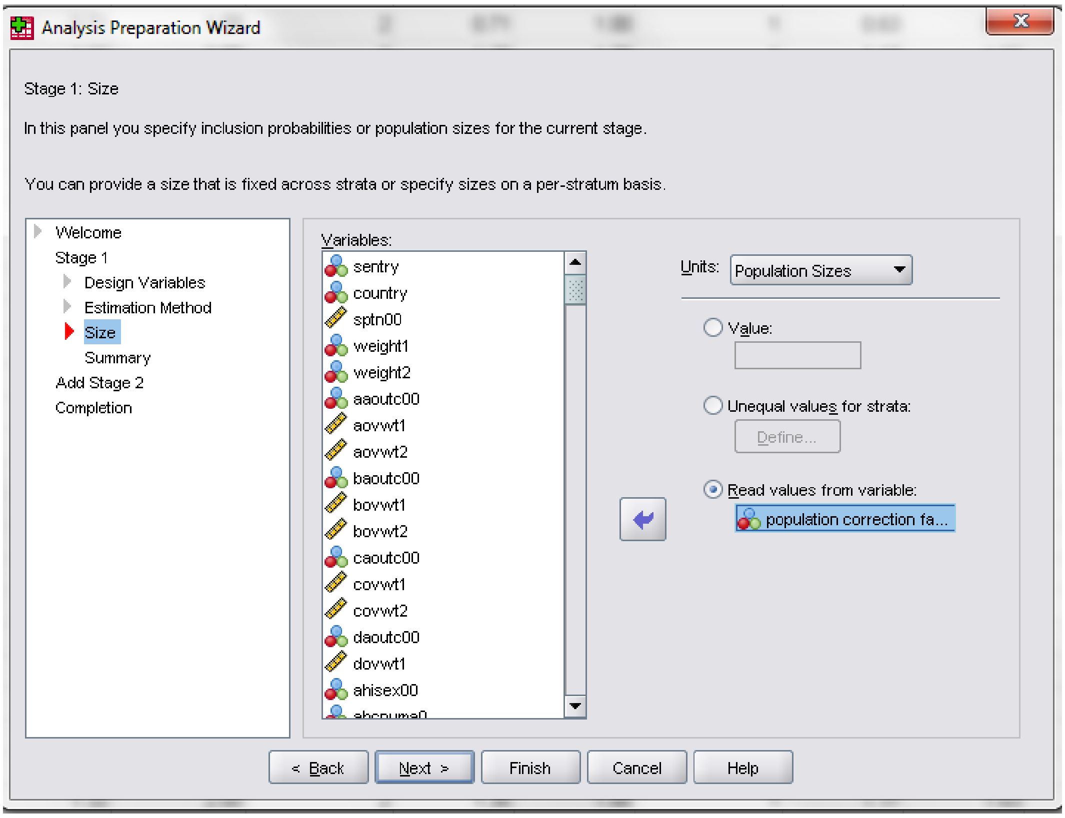
to ***Equal WOR***, then click ***Next***. (Notice the grayed out but selected FPC)



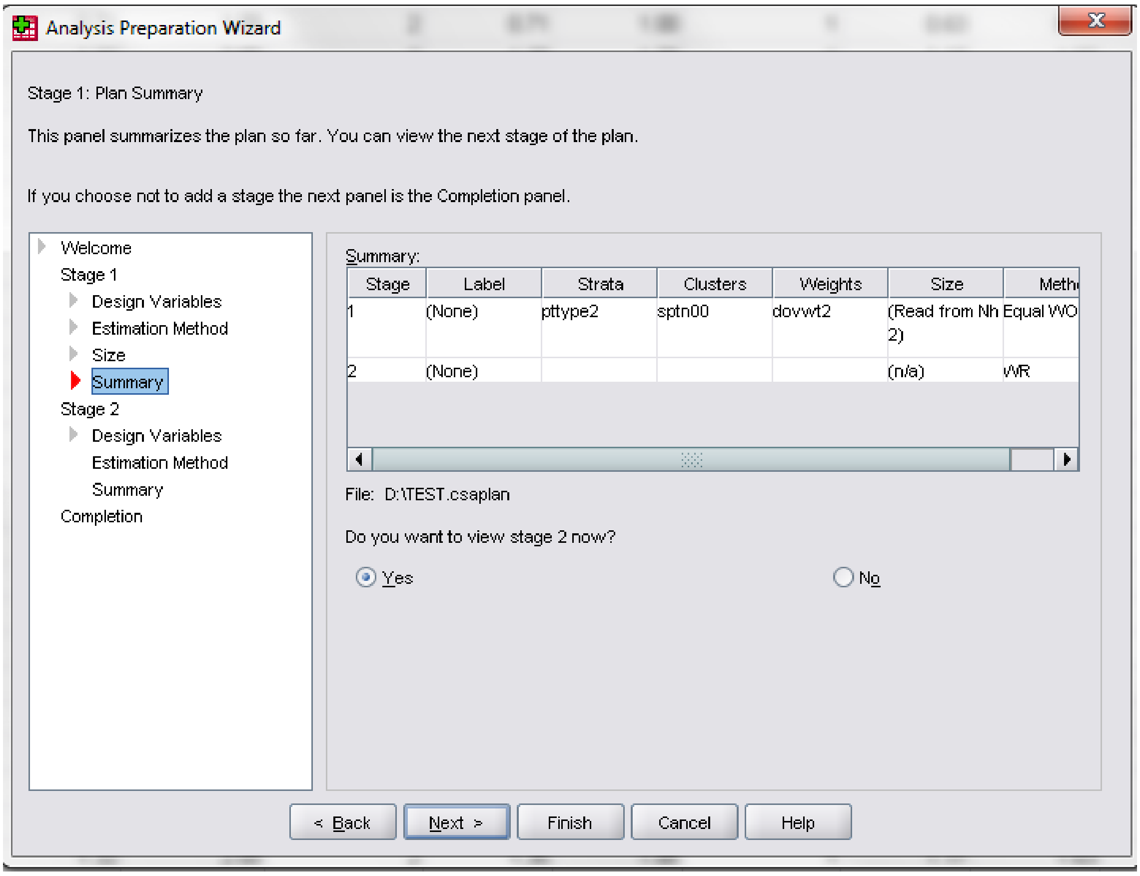
The ***Units*** will be set to ***Inclusion Probabilities***. You want to change this…



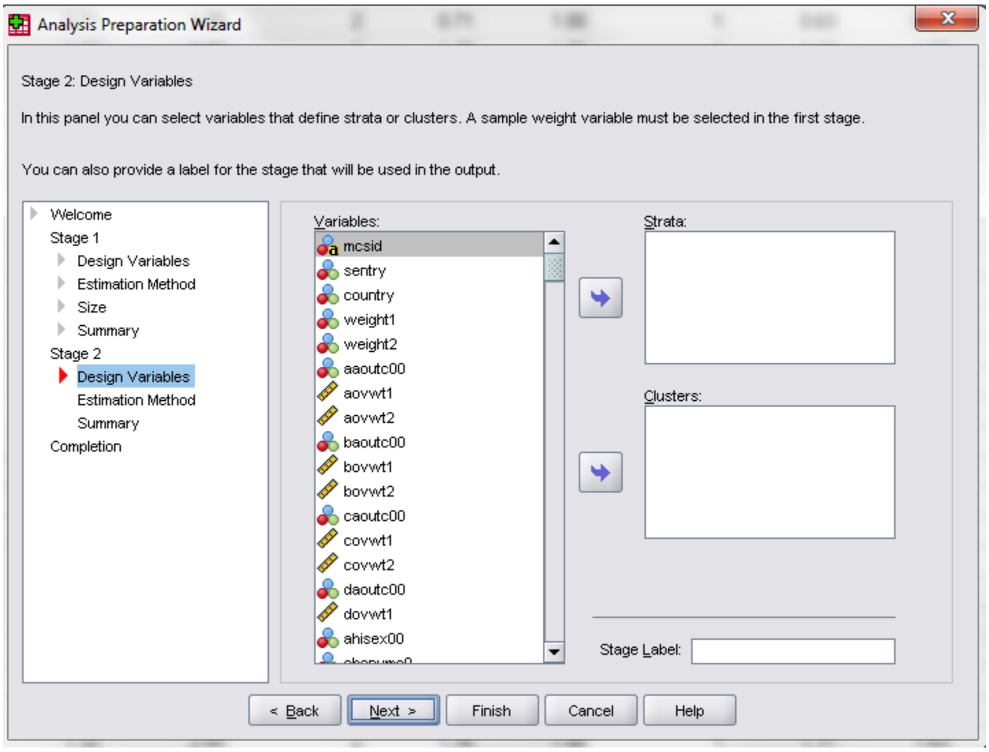
…to ***Population Sizes***. Once you have done that, select ***Read values from variable***, and select **Nh2** as the variable. Click ***Next***.



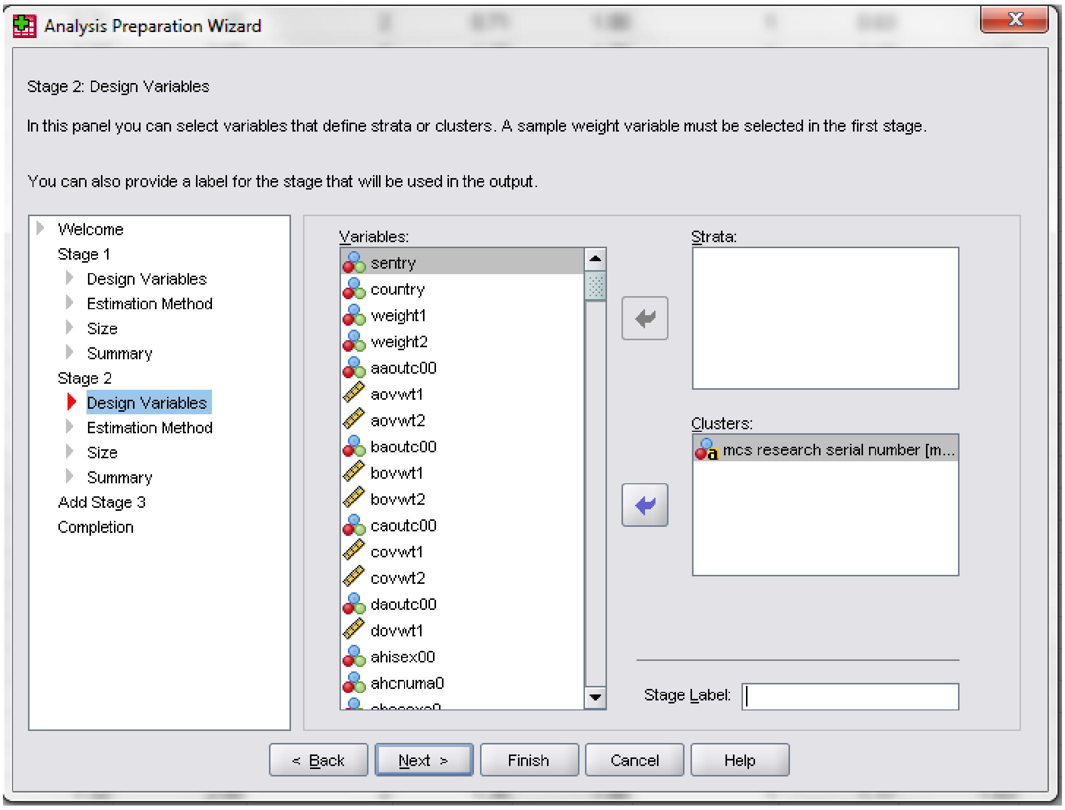
You will then see a summary of what you have set up so far. Make sure the variables in the table are correct, then select ***Yes*** under ***Do you want to view stage 2 now*** and click ***Next***.



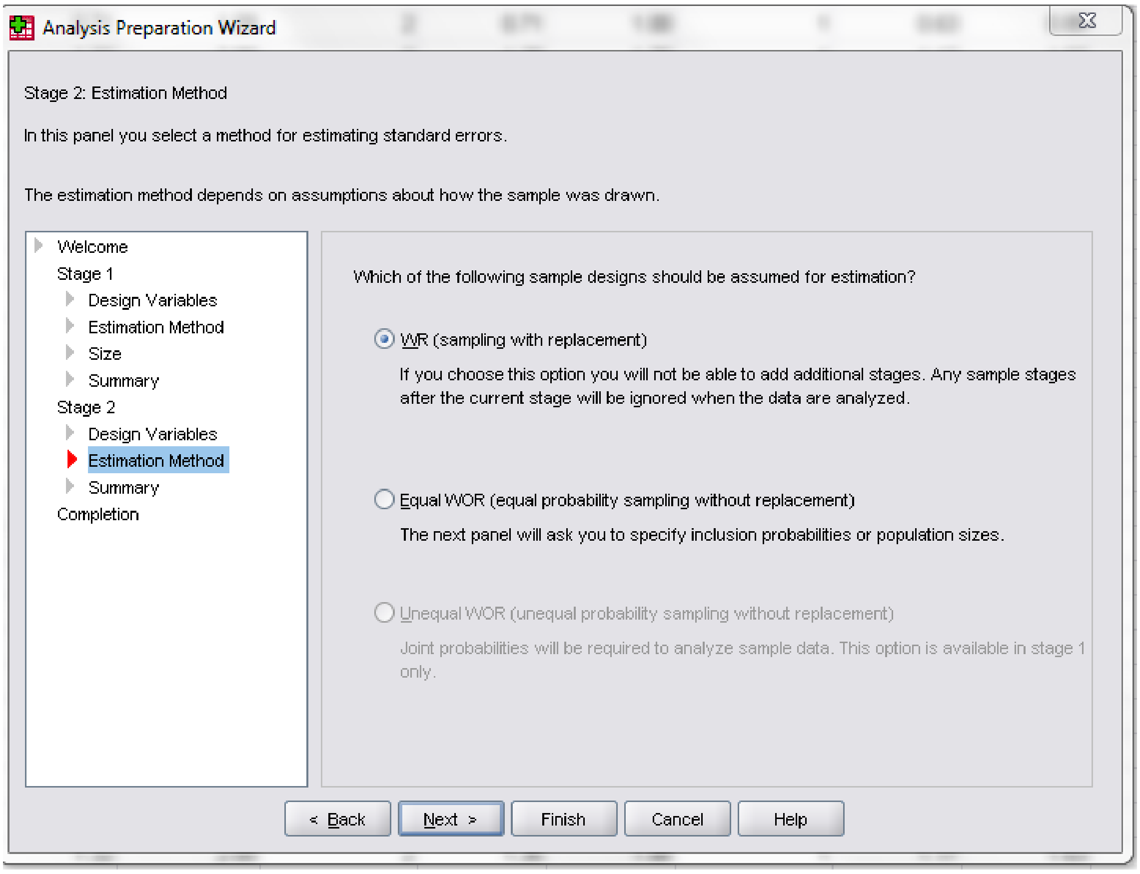
You will see the same dialog box you saw earlier.



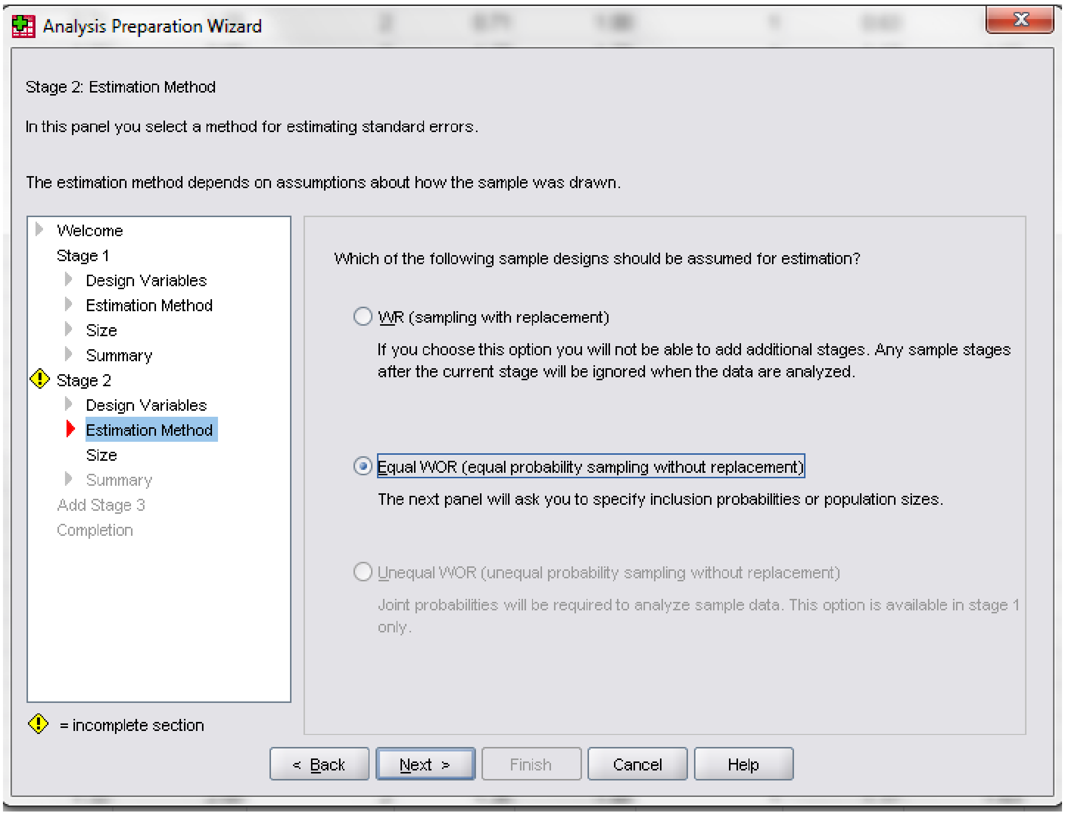
This time you will leave the ***Strata*** box empty and set **mcsid** as the ***Clusters*** variable.



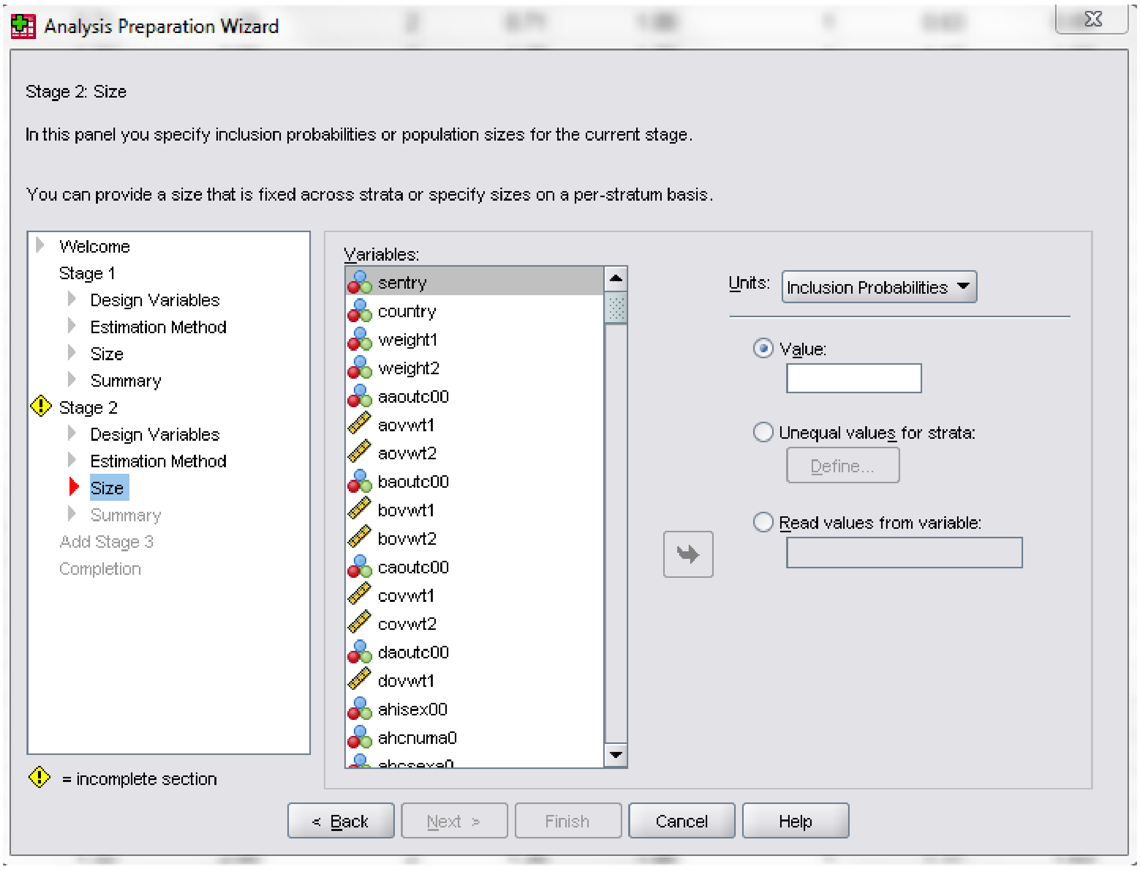
Again, change it from ***WR***…



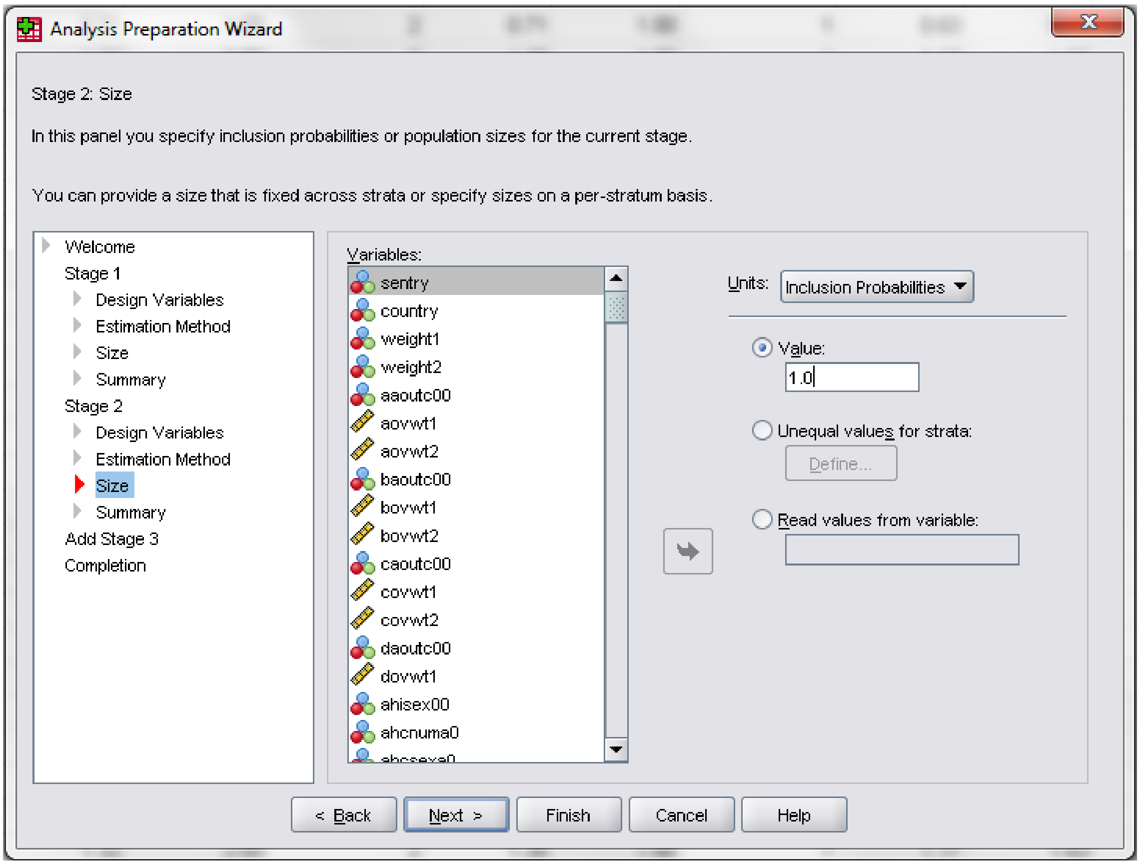
…to ***Equal WOR***, then click ***Next***.



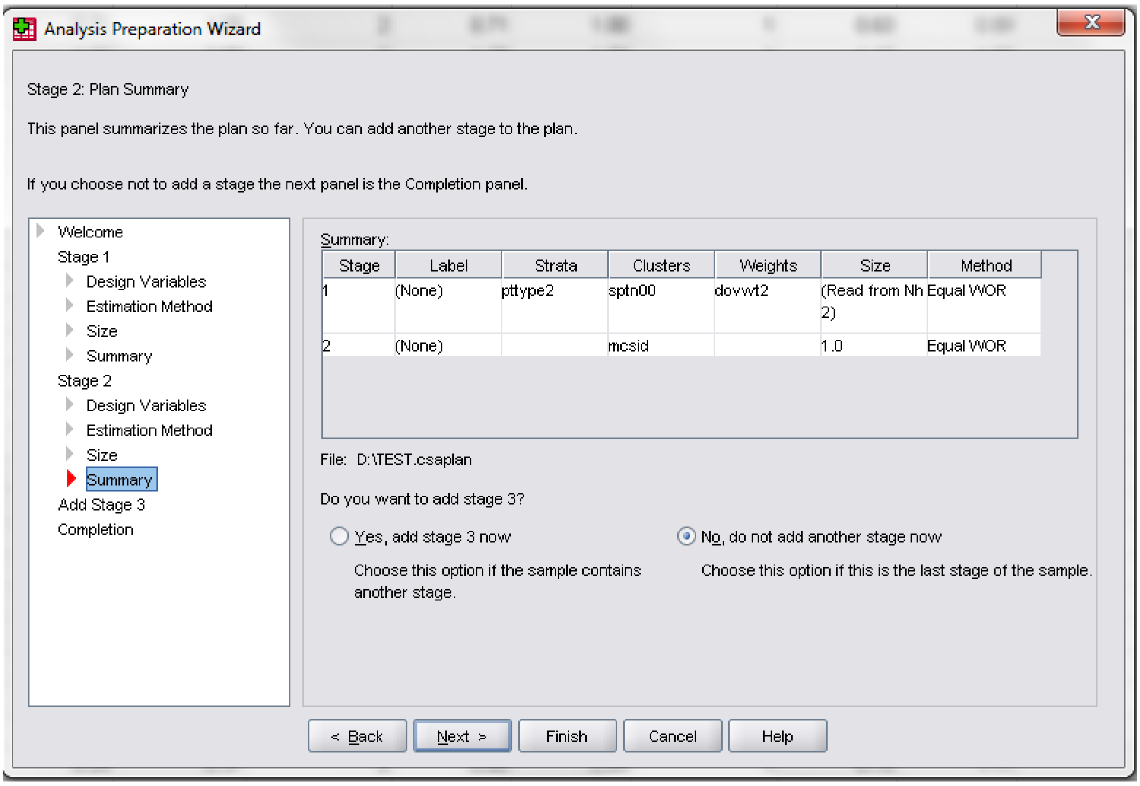
In the next dialog box, leave the ***Units*** as ***Inclusion Probabilities***.



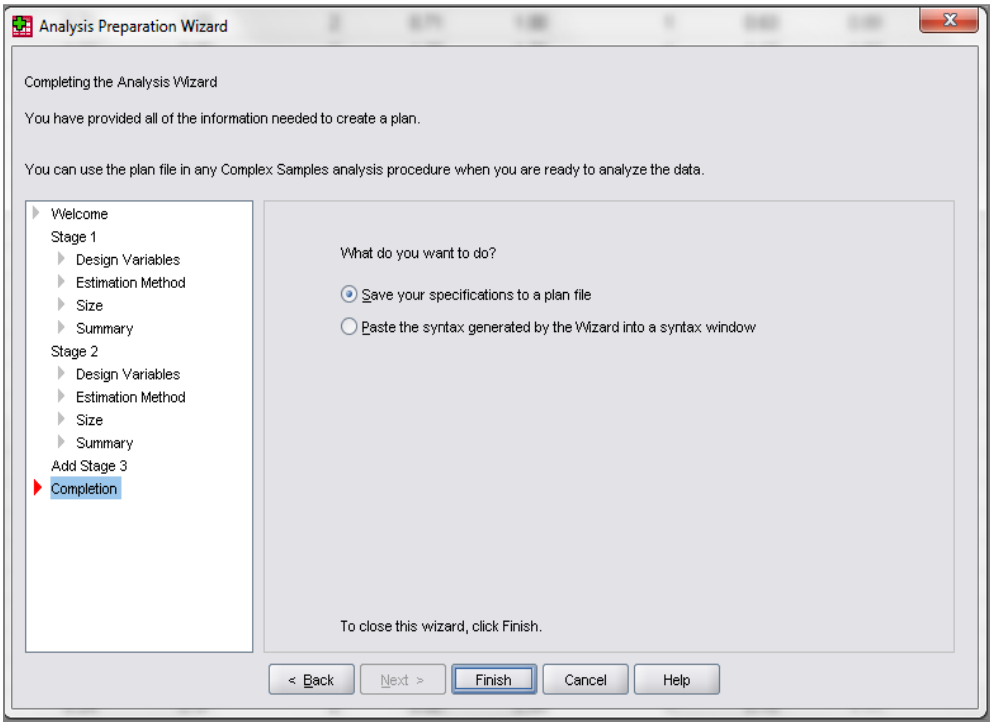
Enter 1.0 as the ***Value***. Click ***Next***.



You will now get a summary of both levels. Select ***No, do not add another stage now***, and click ***Next***.



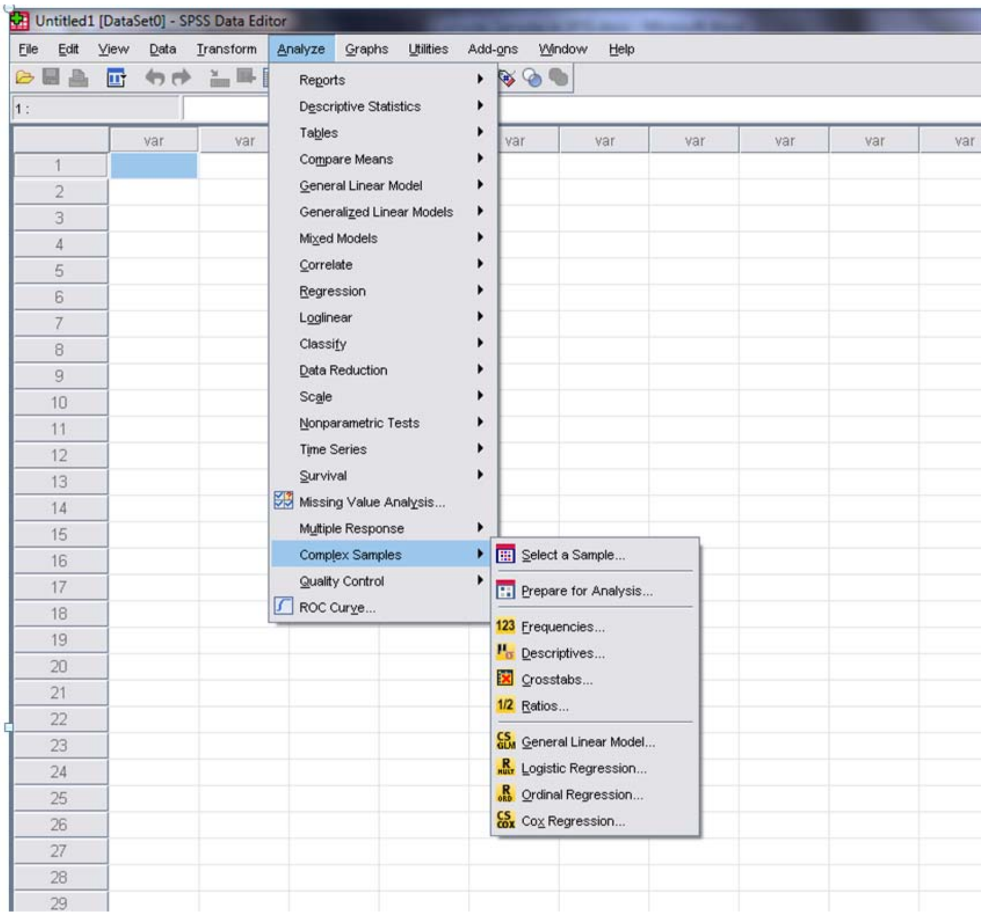
Choose to save your file, then click ***Finish***.



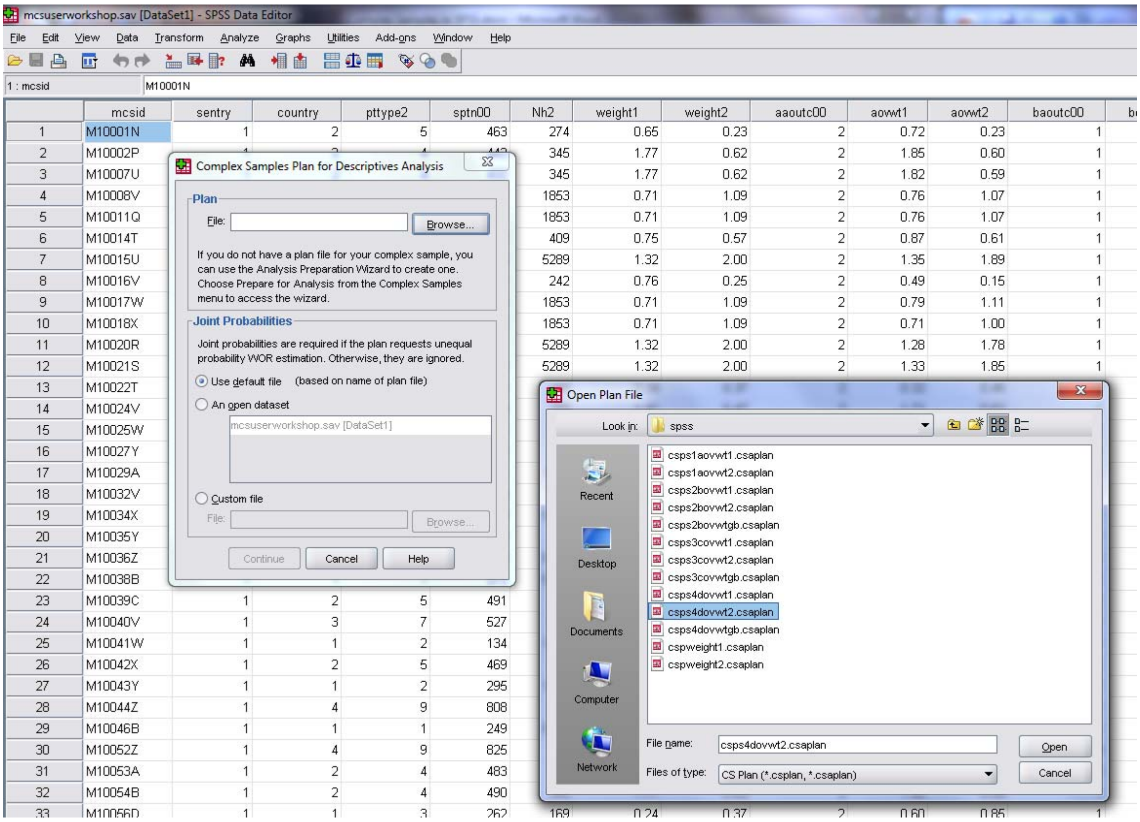
Your CS Plan is now ready to be used.

8. Analyzing data using CS Plan files through menus

Once you have set up your CS Plan file(s) you can analyze data using them. To do this, you need to select analysis tools from the Complex Samples portion of the Analyze menu, as shown below.



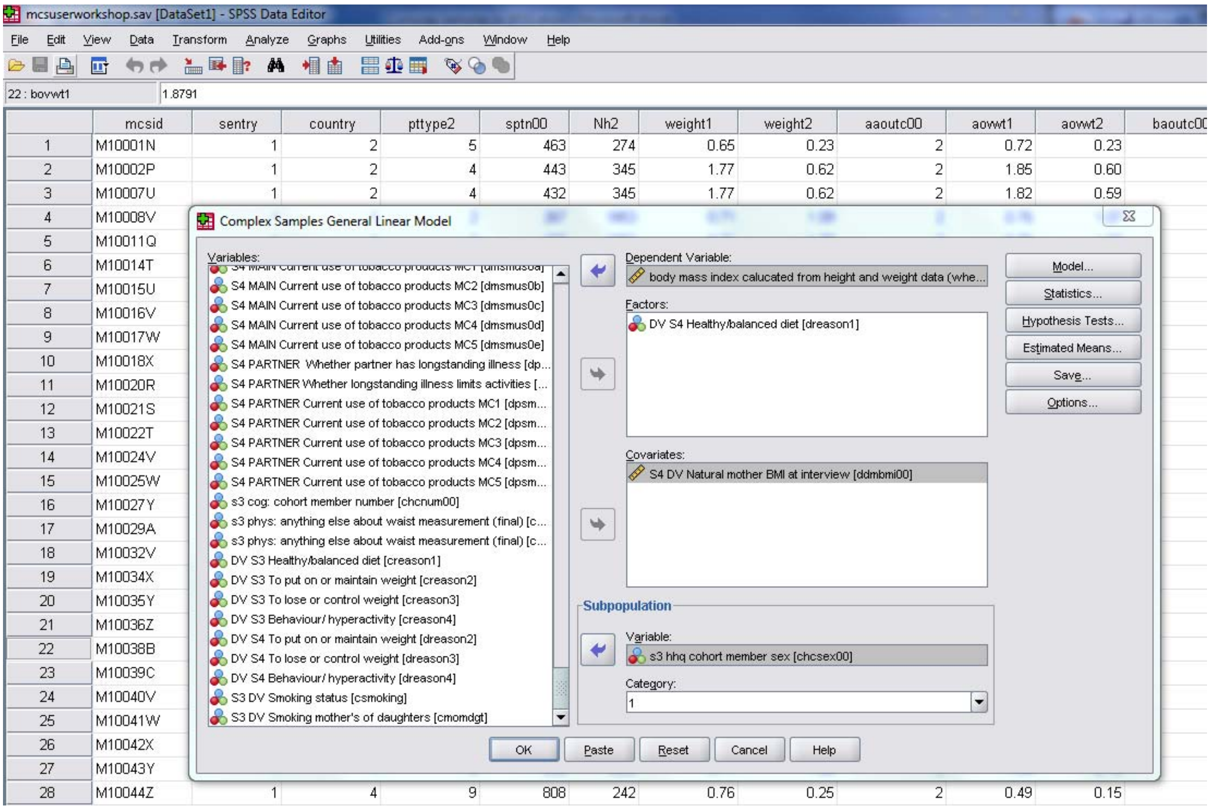
Once you choose your analysis, you will get a dialog box asking for your CS Plan file. Click browse and select the CS Plan file you want to use, based on which weight you need to use for your analysis.



When you have selected your CS Plan file and then clicked ***Continue***, you will be taken to dialog boxes to go through the analysis steps as usual.

9. Performing subgroup analyses with complex samples through menus

When you run analyses using complex samples, you will see a section of the dialog box called subpopulation. Here you can select the variable you want to use and the value of the variable that identifies the cases you want to include. In the example below, we have selected males (category 1 of the cohort member sex variable) as our subpopulation.



Bibliography

Hansen, K (2010). Millennium Cohort Study First, Second, Third and Fourth Surveys: A Guide to the Datasets, Fifth Edition.
http://www.cls.ioe.ac.uk/core/documents/download.asp?id=1260&log_stat=1

Ketende, SC (2010). Millennium Cohort Study Technical Report on Response, Third Edition.
http://www.cls.ioe.ac.uk/core/documents/download.asp?id=1339&log_stat=1

McDonald, JW and Ketende, SC (2010). Nonresponse Weight Adjustments Using Multiple Imputation for the UK Millennium Cohort Study. CLS Working Paper 2010/6. London: Centre for Longitudinal Studies.
http://www.cls.ioe.ac.uk/core/documents/download.asp?id=1406&log_stat=1

Plewis, I (2007). Millennium Cohort Study: Technical Report on Sampling, 4th Edition.
http://www.cls.ioe.ac.uk/core/documents/download.asp?id=875&log_stat=1

Website
www.cls.ioe.ac.uk