National Centre *for* Social Research

The use of sampling weights in the analysis of the 1998 Workplace Employee Relations Survey

Susan Purdon, Kevin Pickering

Prepared for the WERS98 Data Dissemination Service at the National Institute of Economic and Social Research

July 2001

Contents

1 INTRODUCTION

The use of sampling weights in survey analysis has been the subject of a number of academic papers in recent years, but the implications for data users are not widely known. The tendency has been for analysts to ignore sampling weights for regression analyses, the justification being that, although point estimates, such as means and percentages, may be biased if weights are not used, associations between variables will be approximately unbiased, and the interpretation of models, whether or not weighted, will be broadly the same.

For surveys with weights that are largely unrelated to the variables of interest this standpoint may be legitimate. WERS is, however, different. Not only are the weights highly correlated with most of the key survey variables, but also the range of the weights is such that using the weights can significantly affect how models are interpreted.

This paper describes why and when weights are needed for WERS analyses. We start with point estimates, move on to bivariate analyses, and then finish with regression analyses. The intention is to demonstrate that the arguments for using weights for point estimates extend through to the more analytic techniques. Only under very special circumstances can the weights be ignored. The paper concludes with a short review of the available software and an annotated bibliography of books and papers relating to the analysis of complex survey data.

2 WERS98 SAMPLE DESIGN

Given that the need to use weights when analysing WERS data arises because of the complex sample design used, it is useful to start our discussion with a brief reminder of the main features of the survey design. However, full details of the sample design for both the cross-sectional and panel survey components of WERS98 are documented in the WERS Technical Report (Airey et al, 1999).

2.1 Cross-sectional survey sample design

2.1.1 The establishment sample

The establishment sample was selected from the Inter-Departmental Business Register (IDBR) which is maintained by the Office for National Statistics. In selecting the sample for the cross-sectional survey, the ‘population’ of establishments was divided into six size strata (‘size’ being the estimated number of employees as recorded on the IDBR) and, within size strata, by the Standard Industrial Classification major groups (SIC92) (again, as recorded on the IDBR[1]). Within each

[1] Although both size and SIC92 are recorded on the IDBR, the same information was also collected directly from establishments by interviewers as part of the survey. In many cases the IDBR and the interview data correspond, but there are, inevitably, a significant number of differences.

size stratum a roughly equal number of establishments was selected, the actual numbers varying from 362 for establishments with 10-24 employees to 626 for establishments with 200-499 employees. Within each size-group there was some over-sampling of the SIC92 groups: E (electricity, gas and water supply), F (construction), H (hotels and restaurants), J (financial intermediation), and O (other community, social and personal service activities).

The implication of selecting the sample in this way is that the sample has a very different size profile, and a somewhat different SIC92 profile, to the population. Whereas the population of establishments has a considerable skew towards 'small' establishments, the sample distribution has little or no skew by size (Figure 1). And, whereas the population of establishments has a relatively small proportion in SIC92 major group E in particular, the sample proportions are all reasonably large (Figure 2).

Figure 1: Establishment population and selected sample by size

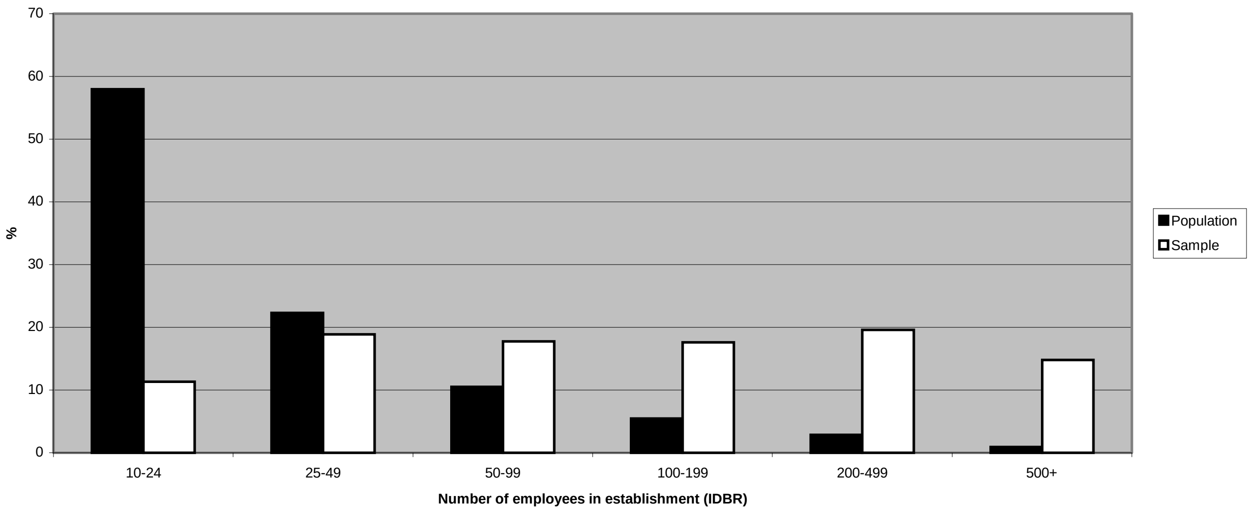
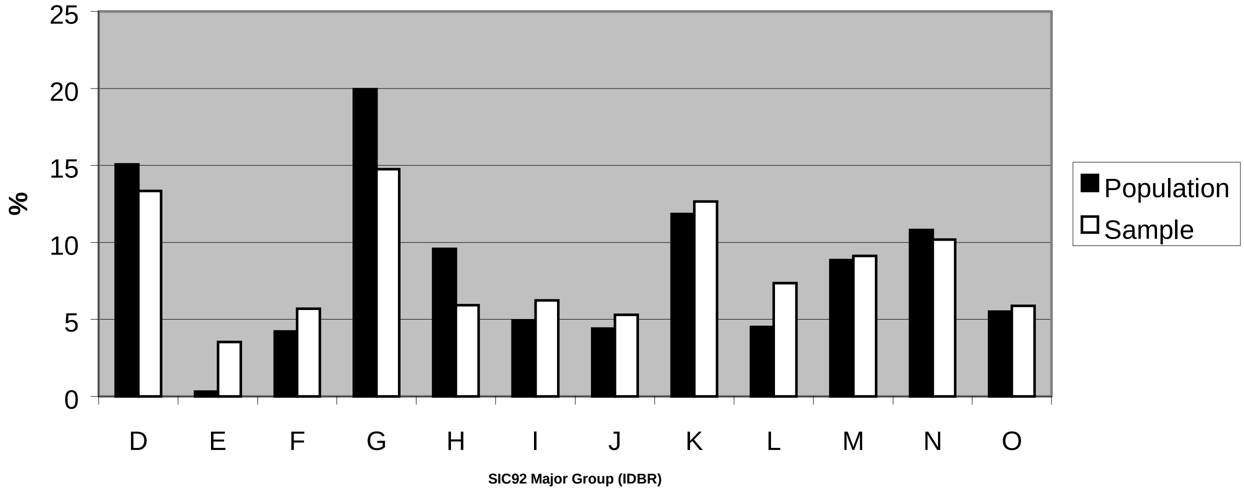


Figure 2: Establishment population and selected sample by sector



The sample design is, thus, a stratified sample, with disproportionate sampling by strata[2]. In total there are 72 strata: 6 size groups by 12 SIC92 major groups. The sampling fractions used per strata vary from 0.0011 in the stratum defined as establishments in SIC92 Group D with 10-24 employees to 0.62 in the stratum defined as establishments in SIC92 Group E with 500 or more employees[3].

The reason for selecting the sample in this way is that it ensures there is an adequate sample size within each size and SIC group to allow for separate reporting and analysis by group. In addition, the over-sampling of large establishments in particular, improves the precision of estimates *about* employees within establishments. (For example, an estimate such as the percentage of all employees in establishments with a recognised trade union is measured with more precision if large establishments are over-sampled.) However, the considerable over-sampling of larger establishments and those in SIC92 major groups E, F, J, H and O, does mean that survey estimates have to be generated by weighting the data to remove the over-sampling biases.

The weights used to create unbiased population estimates from the survey of establishments are included as a survey variable called 'EST_WT'. The weights were calculated as the inverse of the probability of selection for an establishment. Within an individual stratum the weights tend to be fairly constant, but there may be some variation for establishments that proved to be 'non-standard' after selection (for instance, if what was thought to be one establishment at the time of selection turned out to be two). The weights do not include any adjustment for differential non-response, the differences in response rates by establishment size and SIC being fairly small. The details of the calculation of the weights is given in Section 7 of the Technical Report.

---

[2] For interested readers, further information about general sample designs can be found in, for instance, Cochran (1977).

[3] Full details are given in Table 2B of the Technical Report.

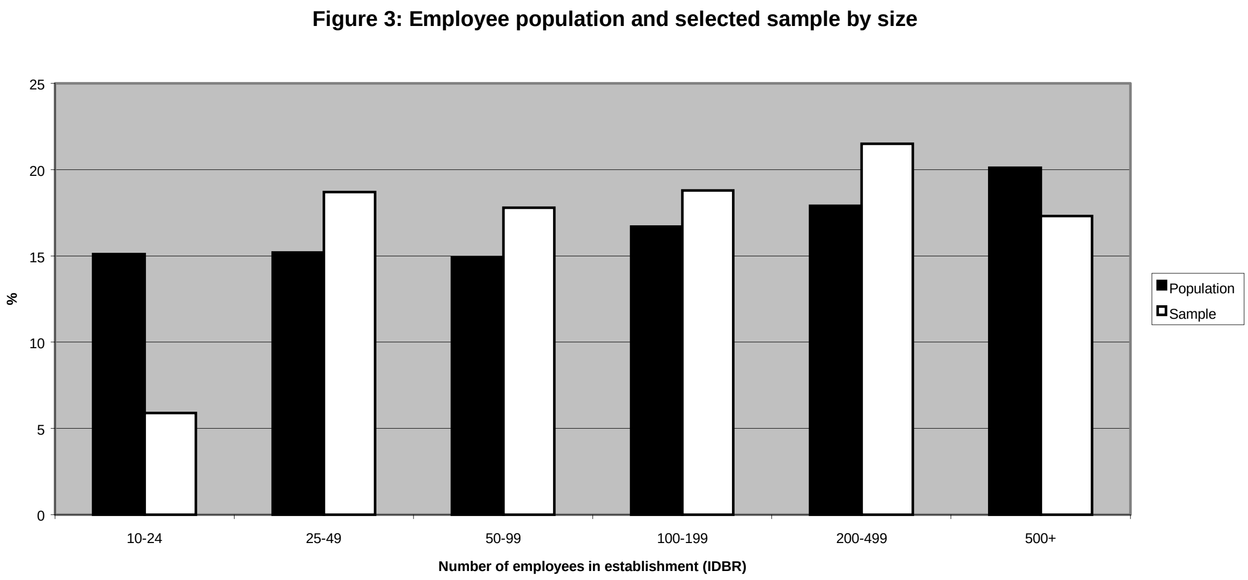
The weights for the establishment survey vary quite considerably from 0.01 to 10.24. Furthermore, this broad range cannot be attributed to a small number of outliers, the 5th and 95th percentiles for the weights being 0.04 and 6.57 respectively. (To put this in context, most household surveys would have weights within a range of about 0.3 to 3.) This large range of weights means that the weights can have a very large impact in analysis.

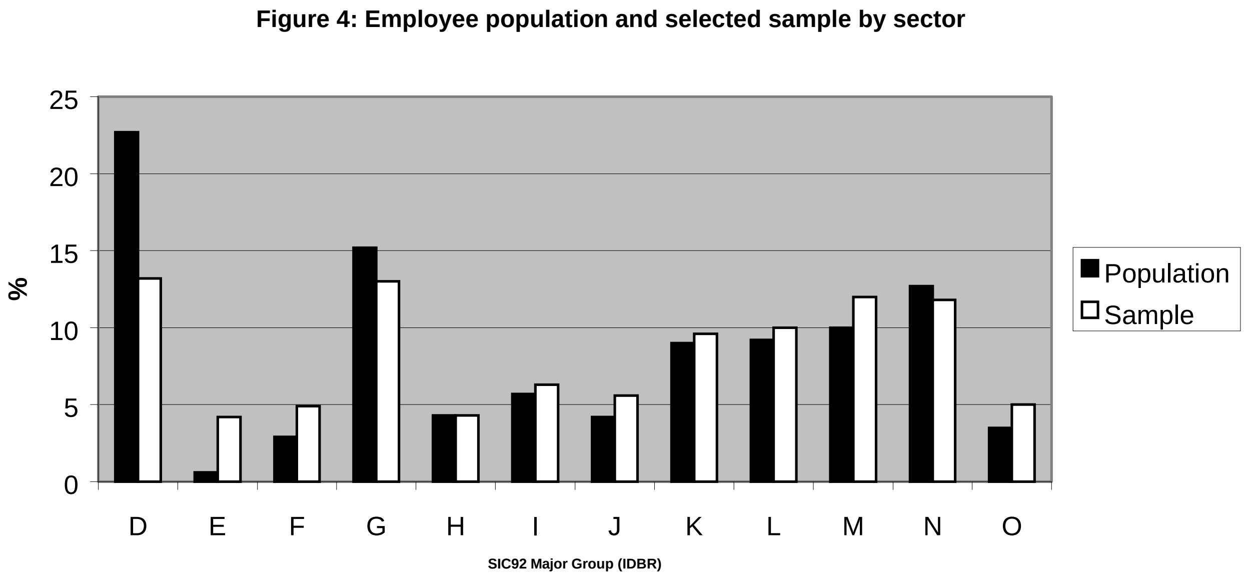
2.1.2 Employee survey sample

Within each establishment selected for (and responding to) the cross-sectional survey, a random sample of 25 employees was selected (or *all* employees were selected in establishments with between 10 and 25 employees). This sample gives a second WERS dataset which has its own set of survey weights (these are included as the variable 'EMPWT_NR').

As with the establishment survey, the weights for the employee sample are calculated as the inverse of the probability of selection for an individual employee but, in this instance, with an additional allowance for differential non-response by gender, part-time/full-time status and broad occupational group. The probability of selection for an employee is calculated as the product of the probability of selection for the establishment and the probability of selection of the employee *within* the establishment. (See Section 7 of the Technical Report for further details.)

The probabilities of selection for employees are, by design, rather less variable for employees than for establishments, primarily because the higher than average probability of selection for the larger establishments is 'balanced' by a lower than average probability of selection of employees within these establishments (and vice versa for smaller establishments). As a consequence the weights for the employee sample range from 0.04 to 17.82 but with 95% of the weights being in the range 0.13 to 3.31. The same point can be demonstrated graphically: Figure 3, which shows the estimated distribution of the population and sample of employees by establishment size, demonstrates that the skew in the establishment sample towards large establishments does not carry through to the employee sample (the population and sample of employees broadly having the same distribution). For completeness, Figure 4 shows the employee population and sample distribution by the SIC92 classification of the establishment.





Selecting the employee sample from *within* the establishments responding to the workplace survey means that the employee sample needs to be handled as a two-stage (or clustered) sample (i.e. the sample of employees is clustered within the sample of establishments). This two-stage sample was stratified at the first stage (i.e. the sampling of establishments), and employees were selected with non-equal probabilities of selection[4].

[4] As was noted earlier, for interested readers, further information about general sample designs can be found in, for instance, Cochran (1977).

2.2 The panel sample

WERS98 also incorporated a panel sample of establishments that were originally selected for the 1990 Workplace Industrial Relations Survey. The 1990 sample itself was selected as a stratified sample, with over-sampling of large establishments and some under-sampling of those in the public administration sector. The 1998 panel was selected as a stratified random sample of 'survivors' from this earlier survey so that the 1990 over and under sampling was automatically incorporated into the panel sample. (The stratifiers in 1998 were based on *interviewer collected* data on establishment size in 1990.)

The sampling weights for the panel survey sample allow for the over and under sampling in 1990 and for some response rate differences in 1998, namely a higher than average response rate amongst public corporations and nationalised industries and establishments with union-related limits on the organisation of work. (See Section 7 of the Technical Report for details.)

This paper does not use the panel survey data to illustrate the use of the sampling weights in analysis. However, the issues for the panel survey and the cross-sectional surveys are the same, and all conclusions drawn about the cross-sectional surveys also apply to the panel survey.

3 THE EFFECT OF THE WEIGHTS AND COMPLEX SAMPLE DESIGN ON POINT ESTIMATES

In this section we consider the impact of the weights and the complex sample design on point estimates. Bivariate and regression analysis are dealt with in Sections 4 and 5 respectively.

Put quite simply, WERS data are weighted to allow for unbiased population estimates to be derived. In other words, by weighting the data, any bias implicitly introduced by the sample design will be removed.

The weights cover one aspect of the sample design, namely the unequal probabilities of selection and some non-response adjustment. The weights, together with the other elements of the design, namely the stratification and the clustering of the employee sample within establishments, all have an impact on the standard errors of estimates derived from the survey data.

These two aspects: weighting to give unbiased estimates, and the impact of the weighting and sample design on standard errors, are discussed in turn below.

3.1 Unbiased estimation

For reasons that are probably by now apparent, for point estimates, such as percentages and means, it is almost always essential that WERS data be weighted when attempting to make inferences about the population. If the data are not weighted then estimates for establishments in particular will reflect the overrepresentation of the larger establishments in the sample (as was discussed in Section 2.1.1).

The weights can make a quite considerable difference to WERS estimates. For example, if we calculate the percentage of establishments with unions recognised for collective bargaining purposes, the unweighted estimate is 56% and the weighted estimate is 39%. Most of the difference between these two figures arises because of the relationship between the survey estimate and establishment size. The table below shows the unweighted percentages by size (as recorded on the IDBR). There is a very strong relationship between size of establishment and union recognition. In creating an unweighted 'all establishment' percentage, the high percentages for the large size categories enter into the estimate in proportion to their *sample* numbers. In the weighted estimate, these percentages are weighted in proportion to the *population* numbers. Because the population is much more skewed towards small establishments than the sample (see Figure 1), the weighted 'all establishment' figure is much closer to the figure for the smaller establishments.

	Number of employees in establishment (IDBR)					
	10-24	25-49	50-99	100-199	200-499	500+
Percentage of workplaces recognising a union	32.2%	42.0%	49.6%	56.4%	66.0%	76.4%

The weights typically have a much smaller impact on estimates of employees because, as noted earlier, the weights for this sample are much less strongly correlated with establishment size. An example of a fairly large difference for employees is the percentage of employees in the SOC major group 'clerical or secretarial' where the unweighted estimate is 23% and the weighted estimate is 18%. A more typical example is the percentage of employees stating that the relationship between managers and employees in their establishment is very good, where the unweighted and weighted estimates are both 15%.

Nevertheless, given that the weights *can* make a difference to employee estimates, we suggest that the weights should always be used to ensure approximate unbiasedness, unless there is good reason to do otherwise.

Arguably the *only* occasion when weights should not be used to present point estimates is where estimates are intended as statements about the sample rather than the population. For example the statement '48% of establishments *in the survey* had fewer than 100 employees' is a statement about the sample rather than the population, so it would be correct to use unweighted data to make this estimate.

In some instances it *might* be argued that the difference between weighted estimates and unweighted estimates is so small that the unweighted estimates can legitimately be presented. This might arise for

- many estimates based on the employee sample;
- estimates based on the establishment sample for variables for which there is very little association with size or SIC;
- sub-group estimates (in particular sub-groups based on the IDBR sampling strata) where the weights are fairly uniform.

However, it should be noted that in making a decision on whether or not to use the weights for point estimates, the effect of weighting on each estimate would have to be checked (by running both the weighted and unweighted estimates) and a decision made about the necessity of weighting each time. This seems to be neither a practical not a theoretically sound approach. In particular, it would be undesirable to have both weighted and unweighted estimates in a single paper.

Our strong recommendation therefore is that weights should always be used to generate point estimates.

3.2 Calculation of standard errors

Relative to a standard error from a simple random sample, the standard errors for WERS estimates are affected by three elements of the sample design[5]:
1. the weighting;
2. the stratification by size and SIC92;
3. the clustering of the employee sample within the establishment sample.

Standard errors that take these elements into account are referred to in this paper as ‘complex standard errors’, although the word ‘complex’ is dropped if it is clear from the context what is meant. Note that the ratio between a complex standard error and a simple random sample standard error for the same estimate is called a design factor (DEFT).

The impact of the three factors listed above on standard errors tends to vary from estimate to estimate, but by and large:
- the effect of weighting is to increase standard errors (relative to simple random sampling standard errors), often quite considerably for estimates based on the establishment sample;
- the effect of the stratification is to slightly reduce the standard errors;
- the effect of the clustering of the employee sample within establishments is to increase standard errors.

The stratification and clustering are simply a feature of the sample design so affect the standard errors of estimates unambiguously. (In other words, these aspects of the design have to be lived with.) The effect of weighting is more complex in that the analyst can choose whether or not to apply the weights. Weighted and unweighted estimates of the same parameter will have different complex standard errors, the unweighted estimate typically having the lower standard error.

This means there is a often a trade-off between bias reduction and standard error minimisation. To minimise bias it is usually necessary, as has been stressed above, to weight the data. However, this weighting will, simultaneously increase the standard error of the estimate. The table below gives examples of estimates from both the

[5] For background information on how sample design affects standard errors see, for example, Lee et al. (1989) p14-15.

establishment and employee surveys, with and without weights, together with their complex standard errors[6].

Alongside the complex standard errors, the standard error calculated under the assumption of simple random sampling is shown. This is the standard error that would be calculated in packages that do not offer *complex* standard error calculations, such as SAS or SPSS (see Section 6 for more details). These figures are shown to demonstrate how inaccurate the standard error calculations from these packages can be.

	Unweighted data			Weighted data		
	Estimate	Complex[7] standard error	Simple random sampling standard error	Estimate	Complex standard error	Simple random sampling standard error
Establishment data						
Percentage of establishments with a recognised union	55.9	0.89	1.06	38.7	1.74	1.04
Mean percentage non-manual employees	58.6	0.66	0.77	58.7	1.25	0.74
Mean percentage part-time employees	26.1	0.47	0.60	32.0	1.14	0.65
Employee data						
Percentage stating relationship between employees and management 'very good'	14.5	0.3	0.2	15.2	0.5	0.2
Percentage working more than 48 hours per week	13.9	0.4	0.2	12.9	0.5	0.2

Returning to the question of whether or not to weight the data, one common way to assess whether the increase in bias associated with unweighted estimates is compensated for by the decrease in the standard error, is to estimate the mean square error. The mean square error, calculated as the square of the bias plus the square of the complex standard error, is the mean, or expected, difference between the true population figure we are attempting to estimate and the actual survey estimate. If the mean square error for the unweighted estimate is smaller than the mean square error for the weighted estimate, the unweighted estimate is, in some sense, 'better' (in that it is likely to be closer to the true population figure than the weighted estimate). If we make the assumption that the weighted estimates in the above table are unbiased the mean square errors for the estimates in the table above are as shown below. Note that, with the exception of the 'mean percentage non-manual employees', the mean square error for the figures shown, are smaller for the weighted estimate than for the unweighted estimate. This suggests that weighted estimates, even with their increased standard errors, are *almost always* preferable to the unweighted estimates. There are, obviously, exceptions, but we would recommend that the weights are used by default.

[6] Software to calculate complex standard errors is described in Section 6. The analysis in this paper has all been carried out using Stata.

[7] This complex standard error takes into account the stratification and clustering, so is not identical to a standard error under the assumption of simple random sampling.

	Unweighted data			Weighted data		
	Complex standard error squared	Bias squared	Mean square error	Complex standard error squared	Bias squared	Mean square error
Establishment data						
Percentage of establishments with a recognised union	0.79	295.8	296.6	3.03	0	3.03
Mean percentage non-manual employees	0.44	0.01	0.45	1.56	0	1.56
Mean percentage part-time employees	0.22	34.8	35.0	1.30	0	1.30
Employee data						
Percentage stating relationship between employees and management 'very good'	0.09	0.49	0.58	0.25	0	0.25
Percentage working more than 48 hours per week	0.16	1.0	1.16	0.25	0	0.25

3.3 Restricted data file

To calculate complex standard errors for WERS survey data it is necessary to specify

(i) the weights to be applied

(ii) the strata

(iii) for the employee data, the clusters or PSUs (primary sampling units). These are simply the establishments.

Unfortunately, for data confidentiality reasons, the strata membership is not given on the standard WERS dataset. This information is held in a separate file, requiring a separate application form to be completed and returned to the Department of Trade and Industry (the official depositor of the data)[8]. It is anticipated that access restrictions may be lifted in due course. However, at the time of writing, the additional application procedure means that it can take some time to obtain data on strata membership.

Without access to the restricted data file there are several means of computing 'approximately correct' complex standard errors. We have considered four:

(i) Assume the survey sample was not stratified. This will tend to overestimate the standard errors;

(ii) Use the interviewer classification of establishment size (NEMPSIZE) as a proxy stratifier;

(iii) Use interviewer collected data on establishment size (NEMPSIZE) and industry (ASIC) to approximate the IDBR strata;

(iv) Create strata by grouping establishments with a similar establishment weight. (For the standard errors shown below we created a total of 63 groups).

Based on just a few estimates as shown in the table below, the third of these options appears to work best, in that it gives standard errors that are closest to the true (i.e. complex) standard errors. We would therefore recommend that, if the restricted data

[8] Application forms are available from the Data Archive at the University of Essex.

file is not available, 'strata' based on interviewer-collected data on size and industry be used instead[9].

	Estimate	True standard error	Standard error with no stratifiers specified (i)	Standard error with interviewer categorisation of establishment size as stratifier (ii)	Standard error based on strata using interviewer data on size and industry (iii)	Standard error with stratification based on recode of weights (iv)
Establishment data						
Percentage of establishments with a recognised union	38.7	1.74	1.94	1.94	1.77	1.88
Mean percentage non-manual employees	58.7	1.25	1.43	1.42	1.28	1.26
Mean percentage part-time employees	32.0	1.14	1.30	1.30	1.16	1.18
Employee data						
Percentage stating relationship between employees and management 'very good'	15.2	0.45	0.50	0.47	0.45	0.46
Percentage working 40+ hours per week	12.9	0.48	0.51	0.50	0.48	0.49

An alternative approach to estimating complex standard errors would be to calculate a simple random sample standard error and then inflate it by an estimate of the design factor (which is the ratio between the complex standard error and the simple random sample standard error). Design factors for key survey variables are given in the Technical Report. However, this is a very crude approach and is likely to be very inexact in many instances. We do not therefore recommend it, especially for standard errors that are to be presented in published papers. The approach may however be useful for exploratory data analysis in packages such as SPSS or SAS.

4 THE EFFECT OF THE WEIGHTS AND COMPLEX SAMPLE DESIGN IN BIVARIATE ANALYSIS

The arguments for using weights to produce unbiased *point* estimates extend quite naturally to tabular or bivariate analysis. The example we use to illustrate this is shown below, a cross-tabulation of union recognition by whether or not the establishment has been at their current address for at least five years. The data used in the table is restricted to establishments in the private sector. Estimates based on both unweighted and weighted data are shown.

[9] SPSS and Stata syntax files that derive a 'strata' variable equivalent to the one used in option (iii) are available from the WERS98 Data Dissemination Service website at: http://www.niesr.ac.uk/niesr/wers98/

	Unweighted data			Weighted data		
Unions recognised	Less than five years at current address	Five or more years at current address	Total	Less than five years at current address	Five or more years at current address	Total
	%			%		
No	72	59	61	83	79	80
Yes	28	41	39	17	21	20
N of cases	*252*	*1247*	*1499*	*303*	*1328*	*1631*

Different analysts use or analyse bivariate data in different ways:

(i) Tables are often used as a convenient way to present point estimates for subgroups. (In the table above this might be the percentage of establishments recognising a union, for sub-groups based on the number of years at the current address.);

(ii) Secondly, analysts may wish to measure the association between two (or more) variables. (This might be done by calculating an overall index of association for a table, such as lambda, or by comparing pairs of point estimates, using measures such as the odds ratio);

(iii) Thirdly, analysts may wish to test whether there is a significant association between two variables in a table. This is usually done using a chi-squared test.

The survey weights impact on all of these approaches. Taking them in turn:

4.1 Using tables to present point estimates

The table shown above might be used quite simply as a convenient way to display the percentage of establishments recognising a union, firstly for those establishments that have been at their current address for less than five years and, secondly, for those establishments that have been at their current address for five or more years. It is apparent that in this instance the weighting has a very large impact on the estimates. The arguments of the previous section obviously apply here: if unbiased estimates for the population are required then the data should be weighted[10].

4.2 Measures of association for tables

Measures of association are usually intended as inferences about the population rather than as specific to the survey. In other words, if, for a two by two table, lambda is calculated, most analysts will wish to interpret lambda as an *estimate* of the association in the general population. If the measure of association is indeed intended in this way, then questions of unbiasedness again arise.

[10] A possible exception to this rule is where the sub-groups being used are the original IDBR strata. In this instance, the weights within a sub-group will all be roughly the same and the weights will not change the estimates by more than a very small amount.

Whether or not the data should be weighted becomes fairly clear once it is observed that all measures of association are simply combinations of percentages from tables. For example, the odds ratio between any two percentages in a table is written as:

$$\left(\frac{p_1}{100-p_1}\right) \Big/ \left(\frac{p_2}{100-p_2}\right)$$

where p_1 and p_2 are the two percentages being compared.

So, for the table shown above, the odds ratio for union recognition between the two 'number of years at address' groups is:

$$\left(\frac{41}{59}\right) \Big/ \left(\frac{28}{72}\right) = 1.79$$ with unweighted data; and

$$\left(\frac{21}{79}\right) \Big/ \left(\frac{17}{83}\right) = 1.30$$ with weighted data.

In this instance the weights make quite a considerable difference. The unweighted data suggests that the odds of an establishment recognising a union are 1.79 times higher amongst establishments who have been at their current address for five or more years than in establishments who have been at their current address for less than five years. With weighted data the ratio of the odds for the two 'number of years at address' groups is rather more moderate at just 1.30. Since the odds ratio based on the weighted data is calculated using unbiased estimates of the individual percentages it is clear that this version is preferable.

Other measures of association, such as lambda and the coefficient of contingency, can also be expressed as combinations of the individual percentages in tables. The same rule applies: to get unbiased estimates of these measures, the percentages used in their calculation should be unbiased. Hence, weighted data should be used.

4.3 Significance testing

Whether there is a significant association between two categorical variables is usually tested using either a chi-squared or a likelihood ratio test. Both of these tests involve the standard errors of the individual percentages of the table, although this is not immediately apparent in the usual algebraic expressions of the tests (which assume simple random sampling).

As with the standard errors for point estimates, significance tests for tabular data are affected by

- the weighting;
- the stratification by size and SIC92;

- and, for tables based on employee data, the clustering of the employee sample within the establishment sample.

A crude adjustment factor for the simple random sample versions of the chi-squared and likelihood ratio tests is to divide the test statistic achieved (after weighting the data) by the average design effect for the individual percentages in the table. (The design effect is the *square* of the ratio of the true standard error to the simple random sample standard error. I.e. the square of the design factor.) For most WERS data this means dividing the test statistic by a factor of about 2.2. However, this is very crude, and a much better way to proceed is to calculate the tests using a software package that can deal with the complex design. (See Section 6 for more details.)

4.4 Extension to bivariate analyses with continuous variables

Although we have concentrated above on bivariate analyses where the two variables being considered are categorical (so that the data can be presented in percentages), the same conclusions can be drawn for continuous data. In particular, if the data is *not* weighted, measures of association (such as ratios of means and correlation coefficients) will usually be biased, and statistical tests, such as the standard t-test, will be biased.

4.5 Restricted data file

As was noted earlier, for correct standard errors to be calculated, the IDBR strata used for sampling need to be known. The same applies to statistical tests of association. However, we noted in Section 3.3 that, if the restricted data file is not available, then defining strata in terms of the interview data on establishment size and industry gives very good approximations to the true standard error. The same appears to be true for statistical tests of association.

5 THE EFFECT OF THE WEIGHTS AND COMPLEX SAMPLE DESIGN IN REGRESSION ANALYSIS

We turn now to the implications of weights for regression analyses.

Regression analyses are typically used for two reasons:

(i) To describe the relationship between, typically, three or more variables, where one of the variables can be thought as a dependent or outcome variable. This is essentially an extension of the bivariate analysis discussed in the previous section.

(ii) To construct a complete behavioural or predictive model of an outcome variable.

Whether or not the data needs to be weighted depends upon which of these two types of models the analyst is trying to fit. Although the terms are not entirely appropriate we use the term ‘descriptive models’ for the first type of model and ‘prediction models’ for the second.

The distinction between the two approaches is that the independent variables used in the descriptive models will be selected on the basis of their direct interest to the analyst. Other variables that are known or suspected to be related to the outcome variable may be excluded from the model because they do not form part of the hypothesis being tested. An example might be an analysis that looked at the relationship between statements about management/employee relationships and gender, after controlling for employee occupational group, but that took no account of any possible relationship with establishment size.

Predictive regression analysis, in contrast, would consider *all* the variables available to the analyst on the grounds that *any* predictors of the outcome variable should be included. Furthermore, if interaction terms were found to be significant these would also be included.

5.1 Descriptive models

For descriptive models the objective of fitting a regression model can generally be expressed as an attempt to measure the association between two variables (the dependent variable and one independent variable[11]) after controlling for the other independent variables in the model. Thus the model can be thought of as a series of 'controlled' bivariate associations.

We described in Section 4 how unbiased measures of association depend upon the individual components of that association being unbiased, and that to achieve this the data must be weighted. By the same reasoning, regression coefficients in descriptive regression models can only be expected to be unbiased if the data is weighted. Furthermore, if the regression coefficients are to be tested for 'statistical significance' then the complex sample design needs to be taken into account. In almost all instances, software packages that ignore the complex sample design will underestimate the standard errors of regression coefficients. The underestimation, could in some instances, be by a factor of two or more.

To illustrate the effect of weighting on regression coefficients, the table below shows the coefficients from a logistic regression analysis of union recognition by number of years at the current address and establishment size. (The coefficients represent the change in the log odds ratio.[12]) The data is restricted to establishments in the private sector. Note that, using unweighted data there is a statistically significant relationship between 'number of years at address' and union recognition, so the conclusion reached would be that number of years at address and union recognition are associated, even after controlling for establishment size. However, the coefficient for 'number of years at address' using the weighted data is considerably smaller and is no longer significant. Using the weighted data we would conclude that, after controlling for establishment size, there is *no* real evidence for any association between number of years at address and union recognition. Since the weights

[11] In some instances the focus may lie upon more than one independent variable, but the principles are the same.

[12] For background information on logistic regression modelling see, for example, Hosmer D and Lemeshow S (1989) *Applied Logistic Regression*, New York: John Wiley and Sons.

remove the selection biases in the survey data we would trust the conclusion from the weighted analysis rather than the unweighted. Note however, that the weights do increase the standard errors of the coefficients.

	Unweighted data (complex standard error)	Weighted data (complex standard error)
Number of years at current address		
Less than five years (reference category)	0	0
Five or more years	-0.51 (0.16)	-0.27 (0.35)
Size		
25-49 (reference category)	0	0
50-99	0.17 (0.24)	-0.04 (0.30)
100-199	0.66 (0.23)	0.40 (0.27)
200-499	1.18 (0.22)	1.17 (0.26)
500-1000	1.90 (0.22)	1.80 (0.26)
1000+	2.20 (0.24)	2.21 (0.29)
Constant term	-1.98 (0.24)	-1.86 (0.39)

The effect of the weights on regression analyses cannot be predicted very easily. The table below shows a logistic regression of the data from the employee survey. The dependent variable in this instance is the employee rating of the relationship between employees and managers, recoded as a binary variable 'very good and good' versus the rest. The independent variables are gender and establishment size. In contrast to the previous model, in this instance not using the weights *under*estimates the strength of the association between gender and the rating given. Again, we should assume the weighted data model is unbiased. But, again, note that the standard errors are larger in the weighted data model.

	Unweighted data (complex standard error)	Weighted data (complex standard error)
Gender		
Male	0	0
Female	0.27 (0.03)	0.38 (0.04)
Size		
10-24 (reference category)	0	0
25-49	-0.22 (0.10)	-0.23 (0.13)
50-99	-0.36 (0.09)	-0.32 (0.11)
100-199	-0.58 (0.09)	-0.58 (0.11)
200-499	-0.66 (0.09)	-0.72 (0.11)
500+	-0.69 (0.09)	-0.83 (0.12)
Constant term	0.54 (0.08)	0.51 (0.11)

The only occasion when weighting will almost certainly not change the regression coefficients is when separate regression coefficients are fitted for each of the original IDBR sampling strata (of which, recall, there are 72). This is because the weights within a stratum are all more or less constant. But to fit separate regression coefficients for each stratum would give an extremely complex model, and is generally not practical, nor desirable, for descriptive models. (It would involve fitting a very large number of interaction terms in the model in order to capture all the between-strata differences). However, if this more complex modelling strategy is practicable, then the weights *can* be left off the data. The main advantage is then that the standard errors of the coefficients will be smaller, so 'significant associations' are more likely to be found. Furthermore, if it can be demonstrated that the regression coefficients are the same for each strata then the interaction terms can be excluded. In this instance the less complex unweighted model will still give unbiased coefficients.

As a general rule we would recommend that, if unweighted data is to be used for regression modelling, then a parallel weighted model is run and a comparison between the coefficients made. If the coefficients do not differ by more than a small amount between the models then the unweighted data can be used, and advantage taken of the smaller standard errors. If, however, the coefficients *do* differ by a non-negligible amount then the weighted model should be used.

It should be re-emphasised, that using unweighted data does not remove the necessity to calculate standard errors that take into account the stratification and the clustering of the employee sample.

5.2 Prediction models

For models where the primary aim is to use all the variables within the WERS dataset to identify the factors that, in conjunction, predict the dependent variable then it should be possible to use unweighted WERS data and still get unbiased regression coefficients. The argument is, in essence, the same as that described in the last paragraphs of the previous section (Section 5.1). In a full prediction model it would be appropriate to test for interactions between the IDBR sampling strata and the other variables in the model. If they are significant they should be left in the model; if not then they can be excluded. Having identified all of the statistically significant relationships present in the data, weighted and unweighted regression coefficients should be the same. Furthermore, with unweighted data, the standard errors of the coefficients will be smaller. Thus, for prediction models, using unweighted data is probably preferable.

However, once again, we would recommend that, if unweighted data is to be used for regression modelling, then a parallel weighted model is run and a comparison between the coefficients made. If the coefficients do not differ by more than a small amount between the models then the unweighted data can be used, and advantage taken of the smaller standard errors. In a prediction model, if there are differences between the coefficients for the weighted and unweighted data models then this would suggest some misspecification of the model and identifying the source of the misspecification should bring the unweighted and weighted models closer together. The need for a comparison between unweighted and weighted models will be particularly acute if interview data on establishment size and sector is used in place of the original IDBR sampling strata since, within interviewer size and sector 'strata', the weights may be far from constant.

Again, it should be noted that using unweighted data in a prediction model does not remove the necessity to calculate standard errors that taken into account the stratification of the establishment sample and the clustering of the employee sample within the establishment sample.

5.3 Restricted data file

For both the descriptive and prediction models, access will at some point be needed to the restricted data file, if only to calculate the correct standard errors around the regression coefficients. If the restricted data file is not available we suggest that strata based on the interviewer data on establishment size and industry be used instead to create approximate 'IDBR strata' (See Section 3.3). This will give slightly inaccurate standard errors but they should be very close to correct.

6 SOFTWARE

This section gives a short overview of the software options for the analysis of WERS98 data.

6.1 Standard statistical packages

Most data users have access to a standard statistical software package, such as SPSS or SAS. These packages tend to be user-friendly and hence are used considerably by the research community. However, as is probably now apparent, these statistical packages are not designed to analyse surveys using complex sampling designs and they are generally not suitable for such analyses.

SPSS and SAS will both analyse weighted data and so will generate unbiased parameter estimates from WERS98 data (i.e. point estimates, measures of association, and regression coefficients). However the standard errors of these estimates will be incorrect, simply because both packages assume the data is generated using a simple random sample. Furthermore, even if the weights are not used in an analysis, SPSS and SAS will still give incorrect standard errors.

In order to obtain 'correct' estimates of standard errors, more specialist software, such as Stata or WesVar, is required. The main features of these two packages are described below. (Other packages, such as SUDAAN, are also available, but are less well known by the authors.)

6.2 Stata

Stata[13] is a more appropriate statistical tool for analysing complex survey data than SPSS and SAS as it includes a family of 'survey' commands that will generate various unbiased estimates with the 'correct' standard errors. These standard errors are estimated using 'linearization'-based variance estimators (Binder, 1983, Lee et al, 1989).

The 'survey' commands in Stata are fairly simple to use and running analyses is relatively quick (unlike WesVar – see below). Once the user has 'told' Stata which variables define the strata, the PSU's and the probability (sampling) weights, every 'survey' command will assume the same design. The family of 'survey' commands

[13] The latest version of Stata, version 7, costs $995 for a single-user license ($499 for academic users) and is available via their website: http://www.stata.com

covers all the types of analyses described in this paper - point estimates (e.g. means, proportions), bivariate relationships (e.g. cross-tabulation) and multivariate relationships (e.g. regression). The practicalities of using Stata's 'survey' commands with WERS data are outlined in Section 4.5.2. of Forth J and Kirby S (2000) *Guide to the Analysis of the Workplace Employee Relations Survey 1999*, London: WERS98 Data Dissemination Service, NIESR. Available on-line at: http://www.niesr.ac.uk/niesr/wers98/

6.3 WesVar

There are two versions of WesVar currently available – version 2 and version 4. Version 2 is free and can be downloaded from their website[14]. The commercially-available version, Version 4, is a more sophisticated package with additional features and a more user-friendly design[15].

WesVar uses replication methods (Brick et al, 2000, Lee et al, 1989) to estimate standard errors for any parameter of interest. In essence WesVar generates multiple sub-sets or 'replicates' of the survey data. An estimate of the parameter of interest is then calculated for each replication in turn. The variation in these replicate estimates gives an estimate of the complex standard error.

Once WesVar has set up the replication weights, they can be used to obtain standard errors for, in theory, any estimate, e.g. for means, medians, or parameters in a regression model. As the replicates only need to be derived once, they can in theory be generated by the data owners and supplied as part of the main database. The data user does not then need to understand the design of the survey - he or she simply has to apply the replication weights.

One advantage of WesVar over Stata is that, in estimating standard errors, replication methods do not assume that the parameters are from a particular distribution, unlike the 'survey' commands in Stata. This means that replication methods are particularly appropriate when the assumptions made by Stata about the distribution of a parameter are invalid – for example, when performing linear regression with non-normal residuals. In addition, this means that standard errors can be estimated for parameters whose distribution is unknown, such as medians or percentiles.

There are five different replication methods available in WesVar. Of these, the jack-knife method is most suitable for the WERS data. In the jack-knife method, the standard errors are estimated by obtaining estimates of the parameter(s) of interest excluding each PSU (establishment) in turn. The sampling variance (and hence the standard error) of a parameter θ is calculated as a function of the difference between each sub-group estimate and the estimate from the full sample.

Unfortunately, Version 2 of WesVar is not able to analyse WERS data directly. This is because there is a limit of 256 on the number of PSUs that can be defined, whereas there are 1,782 establishments (PSUs) for which there are employee data in WERS

---

[14]. WesVar website address: http://www.westat.com/wesvar

[15] The cost of Version 4 is $495 for a single-user license ($350 for academic users).

and 2,190 in the full establishment dataset. (The limit for Version 4 is 9,999 PSUs.) In theory, it would be possible to set up the replicates outside WesVar, but this would need to be done by someone with experience of replication and would produce less accurate estimates of the standard error than would be obtained from using WesVar 4.

It should be noted that WesVar is very computer intensive. Because of the number of replications that need to be fitted, it can take a considerable time to generate the estimates. For example, to fit a logistic regression with two covariates for the employee dataset took over twenty minutes on a standard PC.

7 ANNOTATED BIBLIOGRAPHY

Sampling methods

Cochran W (1977) *Sampling techniques*, 3rd ed., New York: Wiley.

The classic sampling textbook. Detailed and technical.

Kish L (1965) *Survey sampling,* New York: Wiley.

A more readable book than Cochran. Covers practical sampling issues as well as technical details.

Moser C and Kalton G (1971) *Survey Methods in Social Investigation*, London: Heinemann.

A very good, general text on the methodology of social surveys. Chapters 4-7 discuss sampling issues and a wide range of sample designs.

Millward N (1991) "Sampling establishments for social survey research", *The Statistician*, 40, pp.145-52.

A discussion of the particular issues surrounding the sampling of organizational units, based upon experience of the 1980, 1984 and 1990 Workplace Industrial Relations Surveys. Covers questions of: establishment definition; the choice of sampling frame; the use of role-holders; and questions of access.

The sample design for WERS98

Airey C, Hales J, Hamilton R, Korovessis C, McKernan A and Purdon S (1999) *The Workplace Employee Relations Survey (WERS) 1997-8: Technical Report (Cross-Section and Panel Surveys),* London: National Centre for Social Research.

The essential guide to the methodology of WERS98, including full detail on the design of the survey samples and the calculation of the survey weights. Available on-line from the WERS98 Data Dissemination Service web-site: http://www.niesr.ac.uk/niesr/wers98/

The analysis of complex survey data

Lee E S, Forthofer R and Lorimor R (1989) *Analyzing Complex Survey Data*, Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-071, Newbury Park CA: Sage.

A detailed but accessible discussion of the impact of complex sampling designs on the analysis of survey data. Covers the calculation of weights and methods of variance estimation.

Lohr S (1995) *Sampling: Design and Analysis*, Pacific Grove, CA: Duxbury Press.

Wide-ranging text that includes chapters on variance estimation, cross-tabular analysis and regression analysis. More algebra than Lee et al (1989), but well presented with plenty of examples.

Skinner C, Holt D and Smith T (eds.) (1989) *Analysis of Complex Surveys*, Chichester: John Wiley and Sons.

Lohr calls this “The most complete book to date on doing secondary analyses on complex survey data” (1995: 460). One half discusses weighted (or aggregated) approaches to analysis; the other discusses unweighted (or disaggregated) approaches. Very detailed and technical: only recommended for the more advanced reader.

Standard error estimation in STATA and WesVar

Binder, D. A. 1983. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51: 279-292.

A description of the standard error estimation method used in Stata.

Brick, J., Morganstein, D. & Valliant, R. 2000. Analysis of complex sample data using replication. http://www.westat.com/wesvar/techpapers/ACS-Replication.pdf

A description of the standard error estimation method used in WesVar.