ISER
INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

Exploration of statistical weighting in the Graduate Outcomes Survey: technical report

Peter Lynn and Carla Xena
Institute for Social and Economic Research, University of Essex

July 2021

www.iser.essex.ac.uk

Executive summary

HESA commissioned the Institute for Social and Economic Research (ISER) at the University of Essex to carry out research to ascertain whether application of statistical weighting to the Graduate Outcomes survey data can effectively mitigate the consequences of non-response.

Data from the second year of the survey, relating to graduates from the 2018/19 academic year, were used in the research. The findings were subsequently replicated on data from the first year of the survey (2017/18 graduates) to establish the robustness of the conclusions.

A rich set of auxiliary variables were available for weighting and weights were developed based on a number of different non-response models. The weights were tested to see whether they improved the accuracy of estimates of the proportion of graduates in employment and/or study or in highly skilled employment and/or study.

It was found that weighting – based on any of the alternative models – improved accuracy in only a minority of cases, and that the magnitude of the reduction when it occurred was very small. It was consequently concluded that with these data there is no need to use weighted estimation: the accuracy of estimates does not substantially differ between weighted and unweighted estimates.

Users of the Graduate Outcomes Survey data may be reassured by these conclusions, as they indicate that there is no evidence of substantial non-response bias in the survey data.

1 Background

1.1 The Graduate Outcomes Survey

Graduate Outcomes is an annual survey that aims to collect information from almost all graduates of a particular academic year who studied for a higher education qualification at a UK educational institution, approximately 15 months after the completion of studies. The survey is administered by the Higher Education Statistics Agency (HESA) and supersedes the Destinations of Leavers from Higher Education (DLHE) survey. It primarily collects information about the economic activity of graduates, alongside the nature and location of any employment, study or training. In addition, the survey includes questions about previous employment or training since graduation, subjective well-being, as well as some optional sets of questions.

Data from the survey are used for a range of purposes, including by higher education funding bodies and regulators, by researchers and policymakers, and by higher education providers.

The first year of the survey sampled those who had graduated from the 2017/18 academic year and achieved a response rate in the region of 50%. In order to address concerns about the potential for non-response to bias the findings, research was carried out to assess whether weighting would be beneficial. The conclusion (Nathwani and Bermingham, 2020) was that weighting was not needed, but that further research should be carried out once data from the second year of the survey was available. In 2021, HESA commissioned the Institute for Social and Economic Research (ISER) at the University of Essex to carry out further research and this report presents the findings, based on the survey of 2018/19 graduates.

1.2 Non-response bias

Survey nonresponse will introduce bias to survey estimates if nonrespondents are systematically different from respondents in terms of the parameters being estimated. Specifically, nonresponse error is the difference between the numerical value of a survey estimate and the value that would have been obtained if the survey had achieved a 100% response rate. This difference is therefore caused by the fact that not all sample units participate in the survey. Nonresponse error can of course differ for different estimates from the same survey, so it is hard to make general statements about the level of nonresponse error associated with a particular survey. If we consider, for simplicity, simple design-based estimation (so we assume that no weighting, calibration or other statistical adjustment is applied), then the non-response error associated with an estimate $\hat{Y}$ (a sample statistic y that provides an estimate of a population parameter Y) is the product of two components. The first component is the response rate, i.e. the proportion of sample elements providing data to contribute to the estimate. The second component is the difference between the responding units and the nonresponding units in the sample statistic y. We can see this in the following expression for the nonresponse error:

$$(y_r - y_t) = \left(\frac{n_t - n_r}{n_t}\right)\left(y_r - y_{(t-r)}\right) \qquad (1)$$

where y_r denotes the (observed) statistic y based on the responding sample of size n_r;

y_t denotes the (unobserved) statistic y based on the total sample of size n_t;

$y_{(t-r)}$ denotes the (unobserved) statistic y based on the $(n_t - n_r)$ nonresponding units;

The nonresponse error can be seen to depend both on the (estimate-specific) response rate and on the similarity of responding and nonresponding units in terms of the statistic. It is possible for a low response rate to result in little or no nonresponse error for a particular estimate. This will happen if the responding and nonresponding units are similar. But it is also possible for a high response rate to result in

considerable nonresponse error. This will happen if the nonrespondents are rather distinctive in character, i.e. if they are systematically different in terms of the variable(s) that are used to produce statistic y. Although the risk of nonresponse error increases with increasing nonresponse rates, empirical research has found little or no relationship between response rates and nonresponse error (Groves and Peytcheva, 2008).

Nonresponse error can be tackled in both the data collection phase of a survey and the processing/ analysis phase. The two approaches should not be thought of as alternatives; they are complementary and are typically used in combination. In the data collection phase, the focus should be on maximising response rates amongst sample subgroups with low response propensity in order to improve sample composition. This can be done by targeting resources or attention on population subgroups that are known or expected to be under-represented in the responding sample (Lynn, 2017). Such targeting can only be carried out when subgroup membership is known in advance of data collection, for example based on information from the sampling frame. Targeted designs are a subset of a broader class of survey designs known as adaptive designs, in which design variants can either be built in from the outset or introduced in response to fieldwork progress (Schouten et al, 2017). An adaptive survey design approach is used on the Graduate Outcomes Survey, whereby in the latter stages of each survey fieldwork period greater resource and effort is allocated to cases predicted to have a lower propensity to respond.

In the processing/analysis phase statistical adjustment methods such as forms of weighting or calibration can be used to improve the (weighted) sample composition and thereby reduce nonresponse error (Brick, 2013). Tackling nonresponse, whether in the data collection phase or in the processing/analysis phase, will only reduce nonresponse error to the extent that the adjusted variables (the subgroups targeted for response rate enhancement, or the weighting classes) are correlated with survey estimates. The choice of these variables is therefore particularly important.

The Graduate Outcomes Survey is fortunate to have many variables available from the sampling frame that could potentially be suitable for defining weighting classes. These include several related to the graduate's studies, such as subject, degree type, mode of study (full-time or part-time), provider, level of study (undergraduate, postgraduate taught (PGT), postgraduate research (PGR)) and class of degree awarded, as well as demographic indicators such as sex, age, ethnicity, parental education and the Index of Multiple Deprivation. Other social surveys have found such demographic variables to be helpful for weighting[1].

1.3 Weighting

The idea of weighting is conceptually quite simple. In order to make statistical inference from sample to population, the sample units must represent the population in terms of relevant variables. As there are more units in the population than in the sample, each sample unit must represent at least one, and usually considerably more than one, population unit. The number of population units represented by a sample unit is the weight of that sample unit. If certain types of units are under-represented in the sample, those units will tend to have higher weights than others. In general, there are four reasons why subgroups of population units may be under-represented in a survey sample and hence four reasons to apply weighting: sampling variance, disproportionate sampling, differential nonresponse, and coverage error (Biemer and Christ, 2008). Typically, a weighting scheme will involve three steps: design weighting (to correct for disproportionate sampling), a nonresponse adjustment, and a post-stratification or calibration adjustment (to correct for sampling error and coverage error).

[1] For example, the Millennium Cohort Study (Centre for Longitudinal Studies, 2020), the British Crime Survey (Kantar, 2020) and *Understanding Society* (Understanding Society, 2020).

The Graduate Outcomes Survey is a census of the graduates from each academic year, so no sampling is involved. Furthermore, the frame can be assumed to have near-perfect coverage. Consequently, sampling variance, disproportionate sampling and coverage error do not apply. The only reason that the sample might not represent well the population is nonresponse. Thus, all design weights equal one and the weighting considered here is purely a nonresponse adjustment.

If a set of auxiliary variables $\boldsymbol{x}$ (variables that are available for all units in the gross sample, whether or not they respond to the survey) is associated both with the propensity to respond to the survey and with a survey variable, y, then weighting based on $\boldsymbol{x}$ will reduce nonresponse bias in estimates that involve y. Little and Rubin (2019) defined three scenarios for nonresponse: if data are *missing completely at random* (MCAR) the nonresponse mechanism is unconditionally independent of y; if data are *missing at random* (MAR) the nonresponse mechanism is independent of y, conditional on $\boldsymbol{x}$; if data are not missing at random (NMAR) nonresponse is dependent on y even after conditioning on $\boldsymbol{x}$. Unweighted analysis assumes MCAR whereas weighted analysis (with weighting based on $\boldsymbol{x}$) makes the less strong assumption of MAR. In reality it is unlikely that $\boldsymbol{x}$ will completely define the nonresponse mechanism (NMAR): weighting will remove only the proportion of nonresponse bias that is explained by $\boldsymbol{x}$.

The following simple example, using entirely fictitious data, illustrates how this works:

Imagine a survey that aims to estimate the proportion of people suffering from poor mental health and suppose that the sample is selected from a frame that indicates gender of each person. A sample of 2000 people is selected, 1000 men and 1000 women. Response rate is 51.8% amongst men and 61.9% amongst women. Furthermore, assume that 50 male respondents are classified as being in poor mental health on the basis of their survey responses, as are 30 female respondents. This scenario is summarised in Table 1. If all respondents are given an equal weight of size w in analysis (sometimes referred to as 'unweighted' analysis), theoverall proportion in poor mental health would be estimated as:

$$\hat{\bar{Y}} = \frac{(50\times w)+(30\times w)}{(518\times w)+(619\times w)} = \frac{80}{1{,}137} = 0.0704$$

Now, suppose instead that weights are applied to compensate for the difference in response rates between men and women. Thus, men would get a larger weight (1.931) than women (1.616). The weighted estimate of the overall proportion in poor mental health is:

$$\hat{\bar{Y}} = \frac{(50\times 1.931)+(30\times 1.616)}{(518\times 1.931)+(619\times 1.616)} = 0.0725$$

The effect of the weighting on the estimate can be seen to be modest. Unweighted, we estimate that 7.04% of people have poor mental health, whereas the weighted estimate is 7.25%. In a real survey situation, we would assume that the weighting has reduced the effect of nonresponse bias on the estimate, but we would not know whether this constitutes removing most of the bias, or just a small part of it. That would depend on the extent to which gender explains the non-response bias in mental health. If we suppose that the true proportions with poor mental health in the complete sample of 2,000 are 10% amongst men (100 out of 1,000) and 5% amongst women (50 out of 1,000), then the true sample proportion would be 0.0750 or 7.50%. In that situation, we can now see that the weighting has removed roughly half of the total nonresponse bias. In other words, gender explains about half of the nonresponse bias, but bias remains as within both gender groups people in poor mental health were less likely to respond to the survey.

Table 1 Simple illustration of weighting

	Male	Female	Total
Selected sample	1,000	1,000	2,000
Responding sample	518	619	1,137
Response rate	51.8%	61.9%	
Weight	$= 1.931$	$= 1.616$	
Respondents with poor mental health	50	30	80

Nonresponse adjustment involves estimating the response propensity for each responding case and then deriving the adjustment weight as the reciprocal of the estimated propensity. The propensities are estimated by a statistical model of some kind. The estimates therefore depend on two key features of the model (Lynn, 2005). The first is the set of auxiliary variables in the model; the second is the form of the model.

In this study, logit modelling was used to estimate the response propensities as this is appropriate for a binary outcome variable and allows for inclusion of auxiliary variables of all forms (continuous, ordinal, nominal). *Cell weighting* or *class weighting* is often used for nonresponse adjustment, where the gross sample is divided into a set of comprehensive and mutually exclusive classes and the observed response rate in each class is treated as the estimated response propensity for all cases in the class (Lynn, 1996). If the classes are defined by the cross classification of a set of categorical auxiliary variables $\boldsymbol{x}$, then this is equivalent to logit modelling with the saturated model for $\boldsymbol{x}$. Logit modelling is therefore preferable as it allows selection of variables and interactions based on criteria of statistical significance and consequently tends to lead to more parsimonious models and to produce more efficient estimates. Another alternative way to develop nonresponse adjustment weights is through the use of iterative proportional fitting (raking). The main advantage of raking is that it can be used when the population cross-classification of two or more variables is not known while the marginals are (e.g. if they come from separate sources). This is not the case with the Graduate Outcomes survey. Raking and logit modelling can both handle small marginal category sizes better than complete cross-classification but raking generally results in weaker bias reduction as marginal effects are controlled approximately rather than exactly (Deville et al, 2012).

Another possible alternative to logit modelling would have been to use classification models (random trees) or random forests (Breiman, 2001; Lin and Jeon, 2006). Random trees have the potential to improve precision in one-off applications, while random forests overcome the tendency of trees to overfit and therefore improve precision. But in both cases the likely complex non-hierarchical nature of the resultant model makes it difficult to explain or justify. For the Graduate Outcomes survey, the approach to weighting needs to be transparent and suitable for repeated application each year, so random trees and forests were ruled out on these grounds.

2 Nonresponse on the Graduate Outcomes Survey 2018/19

The overall response rate for the survey of 2018/19 graduates was 52%. Appendix 1 shows that the response rate does not differ greatly by sample subgroup. One of the biggest differences observable is between UK-domiciled and non-UK graduates (57% vs. 38%). There is also a considerable difference in response rate between graduates graduating with different classes of first degree, from 41% of those with a third class award to 62% of those with a first class award. There is also some variation between subject areas, with the extremes being business and administrative studies (43%) and physical sciences (61%).

On the other hand, there was less variation in response rate by provider type, mode of study or country of provider.

Demographic variables do not appear to be strongly associated with propensity to respond to the survey. Response rates varied little by sex, age, ethnic group or disability. The biggest demographic differences were by age for the UK-domiciled population, with response being highest amongst those who were either under 19 (61%) or over 35 (61%) at the time of entry to the course. In contrast, a response rate of 52% was obtained amongst 26-31 year-olds.

3 Methodology

3.1 Overall analysis strategy

The analysis strategy was to develop a number of potential weighting models and then to assess the impact that each of them had on two key survey estimates, both overall and for subgroups. Impact was assessed in terms of the effect on the mean squared error (MSE) of the estimates. Additionally, the impact assessment was replicated on the year 1 (2017/18) survey data in order to provide a check on the robustness of the findings. In the following sections we set out how the models were developed and how the effect on the MSE was assessed. As explained in section 1 above, all models are logit models; the models differ only in terms of the set of auxiliary variables included and the form of those variables.

3.2 Analysis base

For graduates who studied multiple subjects (13% of all graduates), the survey outcome data contained multiple records, one per subject. Each record indicated the proportion of time associated with the subject. For all analysis reported here, graduates were classified to their majority subject. This was therefore achieved by retaining for each graduate only the record associated with the largest proportion of time or, in the case of a tie, a random record.

3.3 Initial selection of variables

As explained in section 1 above, for an auxiliary variable to be useful in determining weights, it must be associated both with the propensity to respond to the survey and with the survey variables, $\{y\}$. We therefore carried out an initial selection of variables, retaining only those that were significantly associated both with response rate and (for respondents) with *at least one* of the following two estimates:

- The proportion of graduates in employment and/or study;
- The proportion of graduates in highly skilled employment and/or study.

The earlier study based on year 1 data (Nathwani and Bermingham, 2020) was only able to use the first of these two y-variables. Recent introduction of the coding of occupation to SOC2020 (Office for National Statistics, 2020) made it possible to extend the current study to also include the measure of highly skilled employment and/or study.

An individual was deemed to have responded to the survey if they had a valid xactivity field entry and formed part of the publication population. The binary proportion in employment and/or study variable was defined in exactly the same way as the previous study that utilised year 1 data (Nathwani and Bermingham, 2020). That is, individuals who were classified as unemployed or conducting some other activity (e.g. travel) were placed in the group 'not in employment and/or study', while those in some form of employment (including voluntary work) and/or study were categorised as being in 'employment and/or study'. To create the highly skilled employment and/or study marker, individuals in employment had to be based in SOC groups 1-3 to be in the 'highly skilled employment and/or study' group. Those who indicated that their main activity was both employment and study had to be in 'highly skilled' work to be in this category (this was the case for the majority of these graduates in both years 1 and 2). Otherwise, they were placed into the 'not in highly skilled employment and/or study' category (unless they were in some form of study).

We identified a set of 17 variables to test. For each, we tested the association with response and with each of the survey variables by means of a bivariate logistic regression model. We chose a rather stringent criteria for statistical significance, given the large size of the sample and the need for effects to be substantial in order for weighting to have a beneficial effect. We retained a variable for analysis only if

P < 0.001 both for the association with survey response and for the association with at least one of the two survey variables. These tests were initially restricted to the UK-domiciled population.

The results of the bivariate tests are summarised in Table 2. It can be seen that all 17 models passed the test and were therefore retained for further analysis. The next step was to examine category coefficients. For ordinal variables, only adjacent categories were compared. For nominal variables, all pairwise comparisons were carried out, with the exception of the variable indicating provider, for which this was not feasible as there were 440 categories (providers). Categories for which coefficients were not significantly different (P < 0.001), either in the nonresponse model or in both of the models of survey variables, were combined (but only with adjacent categories in the case of ordinal variables) in order to improve the prediction of response propensity and hence the precision of weighted estimates. The combination carried out at this stage is summarised in the last two columns of Table 2.

All providers with fewer than 100 graduates were combined with other providers of the same type in the same region prior to the bivariate analysis, such that all groups of providers then had a minimum of 100 graduates.

Subsequent to the bivariate analysis, it was decided to combine level of study and class of degree into a single variable with nine categories: this can be seen in Table 3.

3.4 Model building

We first developed a model of nonresponse for UK-domiciled graduates. All variables which had passed the bivariate tests were entered stepwise into a logistic regression model. Variables were entered in increasing order of the smallest category P-value from the bivariate tests (i.e. most highly significant first) and were retained in the model if at least one coefficient had P < 0.001. Any categories of the variable for which P > 0.001 were combined with one or more other categories. At each step, all coefficients for all variables previously entered were also checked and categories were combined if necessary. The large issued sample size (n=579,430 UK-domiciled graduates and n=213,755 non-UK domiciled graduates) provided statistical power to detect effects that were very small in size, so the value of P=0.001 was used as the threshold for significance tests in order to ensure that detected effects were likely to be substantively meaningful. The final model resulting from this process ("Model 1") is summarised in the left panel of Table 3 below.

Model 1 was then extended to encompass non-UK graduates. Variables for which data were present only for UK-domiciled graduates were dropped (ethnicity, index of multiple deprivation, POLAR quintile, parental education) and a different classification of region of domicile was used. After fitting the model following the same procedures as outlined above for UK-domiciled graduates, the final model is summarised in the right panel of Table 3.

The next step in the modelling was to consider the possible role of interaction terms. Four potential interactions were tested, sex by provider, ethnicity by provider type, age by provider type and sex by provider type. Interaction terms were tested for their contribution to the model by adding them, stepwise, to model 1, separately for UK-domiciled and non-UK graduates. The extended version of model 1, including any retained interaction terms, is referred to as model 2.

Table 2 Summary of bivariate models

Variable	P			Drop?	Merge?	Notes
	M1 (R)	M2 (ES)	M3 (Hi-ES)			
Age on entry (14)	0.000	0.000	0.000	N	Y	From 14 categories to 8
Ethnicity (7)	0.000	0.000	0.000	N	Y	'Other' combined with 'not known'
POLAR quintile (6)	0.000	0.000	0.000	N	Y	Quintiles 3 and 4 combined; Scotland is a separate category
Class of 1st degree (7)	0.000	0.000	0.000	N	N	
Region of domicile (15)	0.000	0.000	0.000	N	N	
Qualification level (5)	0.000	0.000	0.000	N	N	
Region of provider (12)	0.000	0.000	0.000	N	Y	London and SE combined; NE and NW combined
Mode of study (2)	0.000	0.000	0.000	N	N	
Qualifications on entry (11)	0.000	0.000	0.000	N	Y	From 11 to 7 categories
Provider type (4)	0.000	0.000	0.000	N	N	
Index of multiple deprivation (11)	0.000	0.000	0.000	N	Y	From 11 to 5 categories
Parents' education (6)	0.000	0.000	0.000	N	Y	From 6 to 3 categories
Socio-economic classification (7)	0.000	0.000	0.000	N	Y	From 7 to 3 categories
Sex (2)	0.000	0.000	0.000	N	N	
Disability (2)	0.000	0.000	0.000	N	N	
Subject (17)	0.000	0.000	0.000	N	Y	From 17 to 15 categories
Provider (317)	0.000	0.000	0.000	N	Y	From 317 to 162 groups

Notes: In the case of categorical predictor variables, P is the minimum P-value across the set of pairs of categories of the variable. Models: M1(R) predicts response to the survey; M2(ES) predicts being in employment or study at the time of the survey; M3(Hi-ES) predicts being in highly-skilled employment or study. The smallest of the 440 providers were combined within regions to produce a minimum sample size of 100 graduates prior to the bivariate analysis, resulting in 317 grouped providers. The Index of Multiple Deprivation (IMD) is defined differently for each of the four constituent countries of the UK. The variable used in this study indicates the decile of the relevant national index within which the graduate's home residence is located. It therefore indicates the level of deprivation relative to others in the same country rather than relative to the UK as a whole. Sex has two categories as the small proportion of "other" (less than 0.1%) have been combined with the modal category – female – for analysis purposes.

A simplified version of model 1 (referred to as model 3) was also to be tested, as this could potentially have advantages of parsimony and efficiency. To simplify the model, relative to model 1, two changes were to be made:

- Variables were only retained in the model if at least one coefficient had $P < 0.0005$;
- The variables provider, provider type and region of provider replaced by a single combined variable (for both UK and non-UK graduates, though the form of the combined variable differed slightly between the two).

The simplified provider variable was defined as follows. Individual providers were retained as separate categories if they had more than 6,000 UK-domiciled graduates (HE providers) or more than 2,000 (all other categories of providers), but were otherwise grouped within regions and provider types. This resulted in 20 HE providers and 2 other providers remaining as separate categories. The size thresholds are arbitrary but were guided by the results of the empirical combination of categories described earlier. A separate version of the variable was developed for non-UK domiciled graduates, which differed only in that all English FECs were combined into a single category, as were all Northern Ireland FECs, and the Open University was no longer retained as a separate category.

3.5 Assessing effects on mean squared error

For each set of adjustment weights under consideration, weighted estimates of the following two quantities (y-variables) were produced:

- Proportion of graduates in employment and/or study at the time of the survey;
- Proportion of graduates in highly-skilled employment and/or study at the time of the survey.

This was done both for the sample as a whole and separately for subgroups defined by:

- Sex;
- Subject area;
- Provider (grouped);
- Subject area within (grouped) provider.

Additional estimates based on protected characteristics were restricted to UK-domiciled graduates, given concerns over the quality of the data for non-UK graduates for some of these fields:

- Total;
- Sex;
- Ethnicity;
- Disability;
- Age group at entry.

For each weighted estimate ($\hat{y}_w$), a test was carried out to determine whether the estimate was significantly different from the equivalent unweighted estimate ($\hat{y}_u$). If the weighted estimate fell outside the 95% confidence interval for the unweighted estimate it was concluded that weighting had a significant effect. In this instance, the difference between the weighted and unweighted estimates, $\hat{y}_w - \hat{y}_u$ was treated as an estimate of the bias reduction due to weighting. Otherwise, the bias reduction was assumed to be zero.

Subsequently, the effect of weighting on the mean squared error (MSE) of the estimate was estimated as the sum of the difference in the variance of the weighted and unweighted estimates and the square of the estimated bias reduction, thus:

$$\Delta MSE = V(\hat{y}_w) - \mathrm{V}(\hat{y}_u) - \hat{B}^2$$

A set of weights would be adjudged to be beneficial in improving the accuracy of estimates if the MSE were reduced ($\Delta MSE < 0$) for more than 50% of the full sample estimates (both y-variables, for total sample, by sex, subject area, provider and subject area within provider, but limited to analysis bases of at least 100 graduates[2]).

[2] This was done to avoid including estimates of the effects on estimates which are themselves rather imprecise. In such cases the data would not provide sufficient power for it to be possible to identify even quite large bias reduction effects.

4 Results - Nonresponse Models

4.1 Model 1

All sixteen of the variables identified by the bivariate analysis, and after combining level of study and class of degree into a single variable as described in section 3.3 above, contributed significantly to model 1 for UK-domiciled graduates. For several variables, the coefficients for two or more categories were no longer significantly different from each other ($P < 0.001$) in the multivariable context, so these categories were combined. For example, the ethnicity variable was reduced from six categories to four as the estimated effects did not differ between Bangladeshis and Pakistanis, nor between black Africans and those of black Caribbean origin. The final model for UK-domiciled graduates is summarised in the left panel of Table 3.

For the non-UK domiciled graduates, study mode (full time vs part time) and region of domicile were found not to be significant predictors and were dropped from the model. Additionally, several other variables showed fewer significant differences between categories than in the UK-domiciled model, resulting in further combination of categories. This can be seen by comparing the left and right panels of Table 3. For example, region of provider has nine categories for UK-domiciled graduates but only five for non-UK graduates.

4.2 Model 2

None of the four interaction terms tested were significant for either UK-domiciled graduates or non-UK graduates. Consequently, model 2 is identical to model 1 and will not be further considered.

4.3 Model 3

Strengthening the criterion for significance, as set out in section 3.4 above, resulted in the following two variables being dropped from the model, relative to model 1:

- POLAR quintile dropped for UK-domiciled graduates;
- Socio-economic class dropped for non-UK graduates.

This was in addition to the simplification of the provider variable as already described. Model 3 is summarised in Table 4.

4.4 Models 4 and 5

The variable indicating level of study/class of award shows one of the strongest nonresponse gradients (Appendix 1) and is also the strongest predictor of nonresponse in models 1 and 3. The variable indicating provider was also one of the strongest predictors of non-response. It was therefore of interest to test whether considerably more parsimonious models containing only these predictors might produce weights that were similarly effective to those stemming from the more complex models, while potentially having a lesser adverse effect on variance. Thus:

- Model 4: Just one predictor variable, the variable combining level of study and class of degree;
- Model 5: Just two predictor variables, level of study / class of degree, and (grouped) provider.

Table 3 Multivariable logistic model of survey response ("Model 1")

		UK			Non-UK		
Variable	Category	b	se	P	b	se	P
Sex	Female or other	1					
	Male	0.939	0.01	0.000***	1.048	0.01	0.000***
Ethnicity	White	1	.	.			
	Bangladeshi or Pakistani	1.092	0.02	0.000***			
	Indian	1.205	0.02	0.000***			
	Black African or Caribbean	1.198	0.01	0.000***			
Subject	Agriculture, Architecture etc	1	.	.	1	.	.
	Biological sciences	0.886	0.03	0.000***			
	Business & admin studies	0.787	0.02	0.000***	0.670	0.01	0.000***
	Creative arts & design	0.817	0.02	0.000***	0.785	0.02	0.000***
	Education	0.851	0.02	0.000***	0.797	0.02	0.000***
	Engineering & technology	1.102	0.03	0.001*	0.837	0.02	0.000***
	Historical & philosophical studies	0.843	0.03	0.000***			
	Languages	0.813	0.03	0.000***			
	Law	0.737	0.02	0.000***	0.791	0.02	0.000***
	Social studies & Mass comms & documentation	0.838	0.02	0.000***	0.805	0.01	0.000***
	Subjects allied to medicine	0.876	0.03	0.000***			
	Other subjects	1.009	0.03	0.755	0.885	0.02	0.000***
Parental degree	No	1	.	.			
	Yes	1.048	0.01	0.000***			
	DK/NA	0.864	0.01	0.000***			
SEC	Not classified/Never worked	1	.	.	1	.	.
	Higher managerial & professional occupations	1.121	0.01	0.000***			
	Intermediate & Lower supervisory & technical	1.066	0.01	0.000***			
	Lower managerial & professional occupations	1.086	0.01	0.000***			
	All coded				1.228	0.02	0.000***
POLAR Quintile	Quintile 1	1	.	.			
	Quintiles 2-4 & Scotland	1.004	0.01	0.700			
	Quintile 5	0.960	0.01	0.0002**			
	Unknown quintile	0.713	0.03	0.000***			
IMD	Decile 1	1	.	.			
	Deciles 2-5	1.014	0.01	0.219			
	Deciles 6-8	1.054	0.01	0.000***			
	Deciles 9 & 10	1.072	0.01	0.000***			
	Missing	1.125	0.16	0.410			
Disability	Known disability	1			1		
	No known disability	1.177	0.01	0.000***	1.240	0.03	0.000***
Study mode	Full time						
	Part time	0.95	0.01	0.000***			

Region of domicile	East & South East	1	.	.			
	Northern Ireland	0.886	0.02	0.000***			
	Wales	0.906	0.01	0.000***			
	London, SW & Midlands	0.963	0.01	0.000***			
	NE, NW, Y&H, Scotland	0.967	0.01	0.001**			
	Channel Islands & IoM	1.276	0.19	0.101			
Region of provider	East Midlands	1	.	.	1	.	.
	East of England (&W & WM)	0.854	0.02	0.000***	1.252	0.03	0.000***
	London & South East	0.805	0.01	0.000***			
	North & NW	0.837	0.01	0.000***			
	Northern Ireland	0.877	0.03	0.0002**	0.689	0.05	0.000***
	Scotland	0.838	0.02	0.000***			
	South West (& Lond & SE)	0.825	0.02	0.000***	1.207	0.03	0.000***
	Y&H (& North & Scotland)	0.747	0.01	0.000***	0.77	0.03	0.000***
	Wales & West Midlands	0.957	0.01	0.0026*			
Age at entry	18 and under	1	.	.	1	.	.
	19	0.911	0.01	0.000***	0.901	0.02	0.000***
	20	0.825	0.01	0.000***	0.686	0.02	0.000***
	21 to 23	0.875	0.01	0.000***	0.685	0.01	0.000***
	24 to 25	0.848	0.01	0.000***	0.795	0.02	0.000***
	26 to 31	0.849	0.01	0.000***	0.930	0.02	0.0006**
	32 to 35	0.974	0.02	0.12	1.052	0.03	0.10
	Over 36	1.338	0.02	0.000***	1.233	0.04	0.000***
	Age not known	0.528	0.03	0.000***	0.531	0.1	0.000**
Provider type	Alternative Provider	1	.	.	1	.	.
	HE Provider	1.06	0.02	0.013	0.807	0.03	0.000***
	English / N Ireland FEC	1	.	.	1	.	.
Qualification and class	PGR (& UG 1st class)	1	.	.	1	.	.
	PGT	0.795	0.01	0.000***	0.785	0.01	0.000***
	UG: 1st class	1.134	0.03	0.000***			
	UG: upper 2nd class	0.899	0.02	0.000***	0.721	0.01	0.000***
	UG: 1ower 2nd class	0.694	0.02	0.000***	0.524	0.01	0.000***
	UG: 3rd class	0.541	0.02	0.000***	0.458	0.02	0.000***
	UG: unclassified	0.697	0.02	0.000***	0.741	0.03	0.000***
	Oth UG: others	0.513	0.01	0.000***	0.491	0.01	0.000***
	Missing	1	.	.	1	.	.
Qualifications on entry	First degree (& any)	1	.	.	1	.	.
	Level 3 qualification	0.890	0.01	0.000***			
	No formal qual/NA/DK/Oth	0.861	0.01	0.000***	0.773	0.01	0.000***
	PGCE	1.126	0.04	0.0005**			
	Postgraduate (exc. PGCE)	0.930	0.01	0.000***			
	Quals at Level 2 and below	0.898	0.03	0.0003**			
	Missing data	1	.	.	1	.	.
Constant		2.237	0.11	0***	1.5	0.06	0.000***
Pseudo R^2		0.0265			0.0300		
Observations		579,430			213,755		

Notes: categories in parentheses apply to non-UK domiciled graduates. Estimates for the variable 'provider' have been omitted from the table (82 categories for UK-domiciled and 48 categories for non-UK domiciled). The subject grouping "Agriculture,

Architecture etc" consists, for UK-domiciled graduates, solely of Agriculture and related subjects and for non-UK graduates of Agriculture and related subjects, Architecture, building and planning, Biological sciences, Combined subjects, Computer science, Medicine and dentistry, Subjects allied to medicine, and Veterinary science. The subject grouping "Other subjects" consists, for UK-domiciled graduates, of Architecture, building and planning, Combined subjects, Computer science, Mathematical sciences, Medicine and dentistry, Physical sciences, and Veterinary science and for non-UK graduates of Historical and philosophical studies, Languages, Mathematical sciences, and Physical sciences. Socio-economic class (SEC) is included in model 1 for non-UK domiciled graduates as it meets the empirical criteria for predictive power, even though the measure itself is UK-specific. The final model distinguishes only between those with substantive categorisations and those without. This may therefore reflect something other than SEC, such as whether the student provided these details and may therefore be acting as a proxy for the circumstances that tend to lead to that happening.

Table 4 Multivariable logistic model of survey response ("Model 3")

		UK			Non-UK		
Variable	Category	b	se	P	b	se	P
Sex	Female or other	1					
	Male	0.939	0.01	0.000***	1.049	0.01	0.000***
Ethnicity	White, Chinese, Other	1	.	.			
	Bangladeshi or Pakistani	1.079	0.01	0.000***			
	Indian	1.194	0.02	0.000***			
	Black	1.181	0.01	0.000***			
Subject	Agriculture & related	1	.	.	1	.	.
	Biological sciences	0.820	0.02	0.000***			
	Business & admin studies	0.725	0.02	0.000***	0.673	0.01	0.000***
	Creative arts & design	0.758	0.02	0.000***	0.772	0.02	0.000***
	Education	0.785	0.02	0.000***	0.810	0.02	0.000***
	Engineering & technology	1.024	0.03	0.403	0.849	0.02	0.000***
	Historical & philosophical studies	0.790	0.02	0.000***			
	Languages	0.754	0.02	0.000***			
	Law	0.703	0.02	0.000***	0.801	0.02	0.000***
	Social studies & Mass comms & documentation	0.776	0.02	0.000***	0.815	0.01	0.000***
	Subjects allied to medicine	0.813	0.02	0.000***			
	Other	0.947	0.03	0.000***			
Parental degree	No	1	.	.			
	Yes	1.047	0.01	0.000***			
	DK/NA	0.877	0.01	0.000***			
SEC	Not classified/Never worked	1	.	.			
	Higher managerial & professional occupations	1.117	0.01	0.000***			
	Intermediate & Lower supervisory & technical	1.064	0.01	0.000***			
	Lower managerial & professional occupations	1.084	0.01	0.000***			
IMD	Decile 1	1	.	.			
	Deciles 2-5	1.013	0.01	0.269			
	Deciles 6-8	1.053	0.01	0.000***			
	Deciles 9 & 10	1.066	0.01	0.000***			
	Missing	0.847	0.12	0.225			
Disability	Known disability	1			1		
	No known disability	1.173	0.01	0.000***	1.217	0.03	0.000***
Study mode	Full time	1					
	Part time	0.957	0.01	0.000***			
Region of domicile	East & South East	1	.	.			
	Northern Ireland	0.868	0.02	0.000***			
	Wales	0.904	0.01	0.000***			
	London, SW & Midlands	0.958	0.01	0.000***			
	NE, NW, Y&H, Scotland	0.957	0.01	0.000***			
	Channel Islands & IoM	1.213	0.18	0.191			

Age at entry	18 and under	1	.	.	1	.	.
	19	0.903	0.01	0.000***	0.901	0.02	0.000***
	20	0.814	0.01	0.000***	0.684	0.02	0.000***
	21 to 23	0.867	0.01	0.000***	0.678	0.01	0.000***
	24 to 25	0.839	0.01	0.000***	0.782	0.02	0.000***
	26 to 31	0.838	0.01	0.000***	0.910	0.02	0.000***
	32 to 35	0.961	0.02	0.020	1.029	0.03	0.356
	Over 36	1.317	0.02	0.000***	1.204	0.04	0.000***
	Age not known	0.575	0.01	0.000***	0.247	0.05	0.000***
Qualification and class	PGR (& UG 1st class)	1	.	.	1	.	.
	PGT	0.789	0.01	0.000***	0.785	0.01	0.000***
	UG: 1st class	1.126	0.03	0.000***			
	UG: upper 2nd class	0.896	0.02	0.000***	0.715	0.01	0.000***
	UG: 1ower 2nd class	0.691	0.02	0.000***	0.518	0.01	0.000***
	UG: 3rd class	0.536	0.02	0.000***	0.459	0.02	0.000***
	UG: unclassified	0.704	0.02	0.000***	0.738	0.03	0.000***
	Oth UG: others	0.510	0.01	0.000***	0.494	0.01	0.000***
	Missing	0.919	0.04	0.051	1	.	.
Qualifications on entry	First degree (or any quals)	1	.	.	1	.	.
	Level 3 qual (including A levels and Highers)	0.873	0.01	0.000***			
	No formal qual/NA/DK/Oth	0.829	0.01	0.000***	0.758	0.01	0.000***
	PGCE	1.118	0.04	0.0010*			
	Postgraduate (exc. PGCE)	0.925	0.01	0.000***			
	Quals at Level 2 and below	0.895	0.03	0.000***			
	Missing data	1	.	.	1	.	.
Constant		2.845	0.12	0.000***	1.452	0.08	0.000***
Pseudo R^2		0.0230			0.0253		
	Observations	579,430			213,755		

Notes: categories in parentheses apply to non-UK domiciled graduates. The subject category "Other" consists, for UK-domiciled graduates, of Architecture, building and planning, Combined subjects, Computer science, Physical and mathematical sciences, Medicine and dentistry, and Veterinary science, and for non-UK graduates, of Historical and philosophical studies, Languages, Mathematical sciences, and Physical sciences. The SEC category "" includes Routine & Semi-routine occupations and Small employers & own account workers.

5 Results: effects of weighting on estimates

In this section, the performance of four alternative sets of adjustment weights is evaluated. In each case, the weights were derived as the inverse probability of responding to the survey, as estimated by a logistic regression model. The four models considered are:

- Model 1: Derived following the analysis protocol and described in section 4.1 above;
- Model 3: Derived following the analysis protocol and described in section 4.3 above;
- Model 4: A model with just one predictor variable, the variable combining level of study and class of degree;
- Model 5: A model with just two predictor variables, level of study / class of degree, and (grouped) provider.

Summary statistics for these four sets of weights are presented in Table 5. It can be seen that models 1 and 3 provide considerably more variation in weights than models 4 and 5. The variation is particularly small for model 4, which should cast doubts on the ability of this set of weights to have much effect on estimates.

Table 5 Summary statistics for each of the four sets of weights

Weighting method	Model 1	Model 3	Model 4	Model 5
Minimum	1.13	1.21	1.54	1.43
Median	1.70	1.70	1.69	1.71
Mean	1.77	1.77	1.77	1.77
Interquartile range	0.337	0.313	0.068	0.226
Maximum	12.45	4.48	2.31	3.10
Skewness	3.98	1.75	1.39	1.95
Kurtosis	63.58	7.55	4.15	7.86

Table 6 presents weighted estimates, standard errors and estimated reduction in MSE for each set of weights and for each of the two y-variables for the total responding sample. These are presented for the whole sample, and for males and females separately. Table 7 presents estimates of the same quantities restricted to UK-domiciled graduates, for the total, and by sex, disability, age on entry and ethnicity.

For the total sample, all four sets of weights reduce the MSE for all three estimates presented in Table 6. The magnitude of the reduction is, however, rather modest. For example, for the model 1 weights the precision improvement for the total sample estimate is equivalent to just 4% of a standard error for y_1 and 6% of a standard error for y_2. These are very small differences by any standards. The relative accuracy improvements with the other three sets of weights are even smaller, as they are for estimates by sex for all four sets of weights.

All but two of the subgroup estimates see an MSE reduction of less than 5% of a standard error and none of the reductions exceed 10% of a standard error. For subgroup estimates by subject area, most weighted estimates have a smaller MSE than the equivalent unweighted estimate (Figures 1 and 4), but the magnitude of these differences in MSE is again very small. For subgroup estimates by provider (Figures 2 and 5) we see a slightly different story as here the majority of weighted estimates have a larger MSE than their unweighted counterparts. And with the exception of just a couple of outliers, those estimates for which the MSE is reduced see it reduced by a very small magnitude. For subgroup estimates by subject area within provider, weighting increases the MSE about as frequently as it reduces it (Figures 3 and 6).

The overall proportions of estimates for which the MSE is reduced is summarised in Table 8. This proportion is dominated by the estimates for providers, as the number of such estimates far exceeds the number of all other estimates combined. Thus, across all estimates for the full sample the best-performing weights are those derived from model 1, with which 22.3% of estimates of the proportion in employment and/or study, and 15.7% of estimates of the proportion in highly skilled employment and/or study, see a reduced MSE. For separate estimates for the UK-domiciled and non-UK populations it is again the model 1 weights that perform best, but for none of these six sets of estimates (2 *y*-variables x 3 populations) do more than 25% of the estimates have a smaller MSE when weighted. Thus, there is a larger number of estimates for which MSE is smaller with an unweighted estimate. Furthermore, we have already seen that the reduction in MSE is very small in magnitude even when it occurs.

As a robustness check, the same sets of estimates were produced, using weights based on the same four models of nonresponse, for respondents to the 2017/18 (year 1) survey. If anything, the weights were slightly less effective than for the 2018/19 survey. The proportion of estimates for which MSE was reduced using the weights from model 1 was 20.0% for the proportion in employment and /or study, and 10.2% for the proportion in highly skilled employment and/or study (compared to 22.3% and 15.7% respectively for 2018/19). The magnitude of the reductions was again very small. It can therefore be concluded that the findings are substantively very similar.

6 Conclusions

It is clear that weighting – based on any of the alternative models tested in this study – reduces the MSE for only a minority of estimates, and that the magnitude of the reduction when it occurs is very small. Despite a rich set of auxiliary variables being available for weighting, it would seem that none of them are sufficiently strongly associated with both the propensity to participate in the survey and the *y*-variables. Indeed, the results of the bivariate analysis of nonresponse presented in section 2 above are quite striking: response propensity varies very little between categories of most of the auxiliary variables.

The corollary of this finding is that with these data there is no need to use weighted estimation: the accuracy of estimates does not substantially differ between weighted and unweighted estimates.

Users of the Graduate Outcomes Survey data may be reassured by these conclusions as they indicate that there is no evidence of substantial non-response bias in the survey data: of the estimates examined, only a minority exhibited empirical evidence of non-response bias that could be explained by auxiliary variables and even in those cases the magnitude of the bias was very small.

It may however be advisable to reassess the nature of non-response bias on the survey and the likely effectiveness of weighting periodically. Survey response mechanisms may change over time. Given the strength of the findings of the current study, and their stability over two years of the survey, it would seem unnecessary to do this every year, but once every several years may be advisable.

Table 6 Estimates, standard errors and ΔMSE for each of 4 sets of weights, by sex

	$\hat{y}_u$	$\hat{y}_{w1}$	se($\hat{y}_{w1}$)	Δmse_{w1}	$\hat{y}_{w3}$	se($\hat{y}_{w3}$)	Δmse_{w3}	$\hat{y}_{w4}$	se($\hat{y}_{w4}$)	Δmse_{w4}	$\hat{y}_{w5}$	se($\hat{y}_{w5}$)	Δmse_{w5}
y_1													
Total	0.88085	0.87603	0.000549	-2.3E-05	0.87637	0.000545	-2.0E-05	0.87663	0.000536	-1.8E-05	0.87609	0.000541	-2.3E-05
Female	0.88079	0.87563	0.000718	-2.7E-05	0.87592	0.000714	-2.4E-05	0.87636	0.000701	-2.0E-05	0.87570	0.000708	-2.6E-05
Male	0.88093	0.87659	0.000850	-1.9E-05	0.87698	0.000842	-1.5E-05	0.87701	0.000831	-1.5E-05	0.87662	0.000839	-1.8E-05
y_2													
Total	0.67720	0.67044	0.000769	-4.6E-05	0.67126	0.000764	-3.5E-05	0.67443	0.000751	-7.6E-06	0.67337	0.000756	-1.5E-05
Female	0.66468	0.65832	0.001013	-4.0E-05	0.65908	0.001008	-3.1E-05	0.66214	0.000991	-6.4E-06	0.66090	0.000997	-1.4E-05
Male	0.69483	0.68721	0.001179	-5.8E-05	0.68812	0.001171	-4.5E-05	0.69153	0.001151	-1.1E-05	0.69064	0.001160	-1.7E-05

Notes: y_1 in employment or study; y_2 in highly-skilled employment or study; w1 indicates the weights derived from model 1, etc; the small numbers who report 'other' sex are here combined with female.

Table 7 Estimates, standard errors and ΔMSE for each of 4 sets of weights, by sex, age, disability and ethnicity (UK-domiciled only)

	$\hat{y}_u$	$\hat{y}_{w1}$	se	Δmse_{w1}	$\hat{y}_{w3}$	se	Δmse_{w3}	$\hat{y}_{w4}$	se	Δmse_{w4}	$\hat{y}_{w5}$	se	Δmse_{w5}
y_1													
Total	0.88422	0.87603	0.000549	-2.3E-05	0.87637	0.000545	-2E-05	0.87663	0.000536	-1.8E-05	0.87609	0.000541	-2.3E-05
Female	0.88488	0.87563	0.000718	-2.7E-05	0.87592	0.000714	-2.4E-05	0.87636	0.000701	-2.0E-05	0.87570	0.000708	-2.6E-05
Male	0.88327	0.87659	0.000850	-1.9E-05	0.87698	0.000842	-1.5E-05	0.87701	0.000831	-1.5E-05	0.87662	0.000839	-1.8E-05
Disability	0.85216	0.84672	0.001564	-2.4E-05	0.84695	0.001555	-2.2E-05	0.84732	0.001548	-1.9E-05	0.84654	0.001560	-2.6E-05
No disability	0.89048	0.88020	0.000585	-2.9E-05	0.88055	0.000581	-2.5E-05	0.88110	0.000570	-2.0E-05	0.88059	0.000576	-2.5E-05
<21 at entry	0.89058	0.86649	0.000828	-4.1E-05	0.86671	0.000825	-3.8E-05	0.86717	0.000817	-3.3E-05	0.86705	0.000821	-3.4E-05
21+ at entry	0.85748	0.88689	0.000748	-1.6E-05	0.88734	0.000743	-1.3E-05	0.88738	0.000727	-1.2E-05	0.88667	0.000736	-1.8E-05
White	0.89724	0.89537	0.000630	-3.5E-06	0.89540	0.000627	-3.2E-06	0.89521	0.000623	-4.1E-06	0.89484	0.000627	-5.7E-06
Indian	0.86928	0.86628	0.003279	6.4E-07	0.86655	0.003270	5.9E-07	0.86693	0.003252	4.6E-07	0.86676	0.003262	5.2E-07
Pakistani	0.80434	0.79727	0.004356	1.6E-06	0.79784	0.004340	1.4E-06	0.79892	0.004296	1.0E-06	0.79881	0.004309	1.1E-06
Bangladeshi	0.79154	0.78441	0.006316	3.3E-06	0.78301	0.006340	3.3E-06	0.78352	0.006280	2.9E-06	0.78297	0.006311	3.3E-06
Chinese	0.84615	0.84293	0.006913	2.9E-06	0.84318	0.006900	2.6E-06	0.84307	0.006855	2.1E-06	0.84352	0.006847	2.0E-06
Black African	0.84605	0.84001	0.002988	-3.6E-05	0.84078	0.002956	6.8E-07	0.84158	0.002931	5.3E-07	0.84135	0.002937	5.7E-07
Blk Caribbean	0.85140	0.84885	0.005540	1.6E-06	0.84935	0.005513	1.4E-06	0.84794	0.005543	1.7E-06	0.84885	0.005517	1.4E-06
Other ethnicity	0.84622	0.85718	0.001121	-2.2E-05	0.84261	0.002212	-1.5E-05	0.84272	0.002193	-6.1E-06	0.84276	0.002197	-1.2E-05
y_2													
Total	0.66084	0.67044	0.000769	-4.6E-05	0.67126	0.000764	-3.5E-05	0.67444	0.000751	-7.6E-06	0.67337	0.000756	-1.5E-05
Female	0.65050	0.65832	0.001013	-4E-05	0.65908	0.001008	-3.1E-05	0.66214	0.000991	-6.4E-06	0.66090	0.000997	-1.4E-05
Male	0.67576	0.68721	0.001179	-5.8E-05	0.68812	0.001171	-4.5E-05	0.69153	0.001151	-1.1E-05	0.69063	0.001160	-1.7E-05
Disability	0.61250	0.60916	0.002094	-0.0001	0.60963	0.002086	-9.3E-05	0.61290	0.002075	-4.0E-05	0.61163	0.002085	-5.8E-05
No disability	0.67028	0.67917	0.000825	-5.5E-05	0.68004	0.000820	-4.3E-05	0.68383	0.000805	-7.7E-06	0.68279	0.000811	-1.5E-05
<21 at entry	0.66108	0.60855	0.001160	-7.3E-05	0.60935	0.001156	-6.0E-05	0.61308	0.001146	-1.6E-05	0.61314	0.001149	-1.6E-05
21+ at entry	0.65986	0.74020	0.001027	-8.4E-05	0.74126	0.001020	-6.5E-05	0.74391	0.000996	-3.0E-05	0.74325	0.001003	-3.7E-05
White	0.67353	0.66317	0.000971	-0.00011	0.66359	0.000966	-0.0001	0.66577	0.000956	-6.0E-05	0.66490	0.000961	-7.4E-05
Indian	0.69323	0.68211	0.004490	-0.00012	0.68415	0.004467	-9.0E-05	0.68679	0.004431	7.0E-07	0.68641	0.004443	8.0E-07
Pakistani	0.57103	0.55487	0.005329	-0.00026	0.55534	0.005318	-0.00027	0.55990	0.005277	-0.00012	0.55940	0.005291	-0.00013
Bangladeshi	0.51152	0.49126	0.007571	-0.00041	0.49254	0.007549	-0.00034	0.49516	0.007509	-0.00027	0.49557	0.007523	-0.00025
Chinese	0.70128	0.69280	0.008757	4.4E-06	0.69420	0.008726	3.7E-06	0.69544	0.008646	2.5E-06	0.69626	0.008647	2.5E-06
Black African	0.58592	0.57087	0.003971	-0.00023	0.57258	0.003952	-0.00019	0.57418	0.003934	-0.00014	0.57524	0.003934	-0.00011
Blk Caribbean	0.58016	0.56917	0.007650	2.6E-06	0.57055	0.007608	2.0E-06	0.57048	0.007587	1.6E-06	0.57216	0.007584	1.6E-06
Other ethnicity	0.63937	0.62495	0.002945	-7.3E-05	0.62728	0.002926	-4.7E-05	0.62971	0.002897	9.8E-08	0.62982	0.002902	-1.0E-05

Notes: y_1 in employment or study; y_2 in highly-skilled employment or study; w1 indicates the weights derived from model 1, etc; the small numbers who report 'other' sex are here combined with female.

Figure 1 **Boxplot of the change in MSE for estimates of the proportion in employment and/or study for each of 19 subject areas, by weighting model**

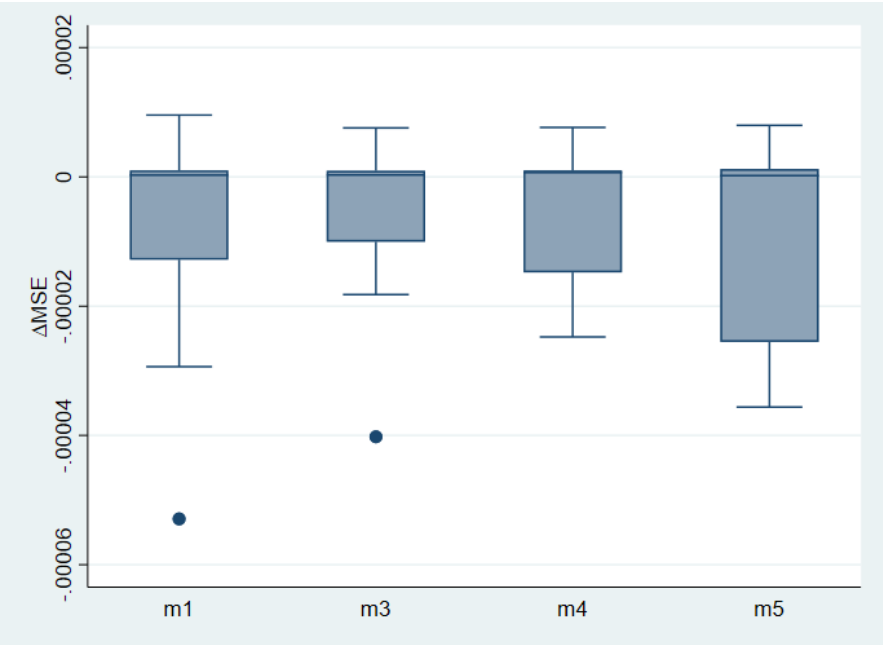

Figure 2 **Boxplot of the change in MSE for estimates of the proportion in employment and/or study for each provider, by weighting model**

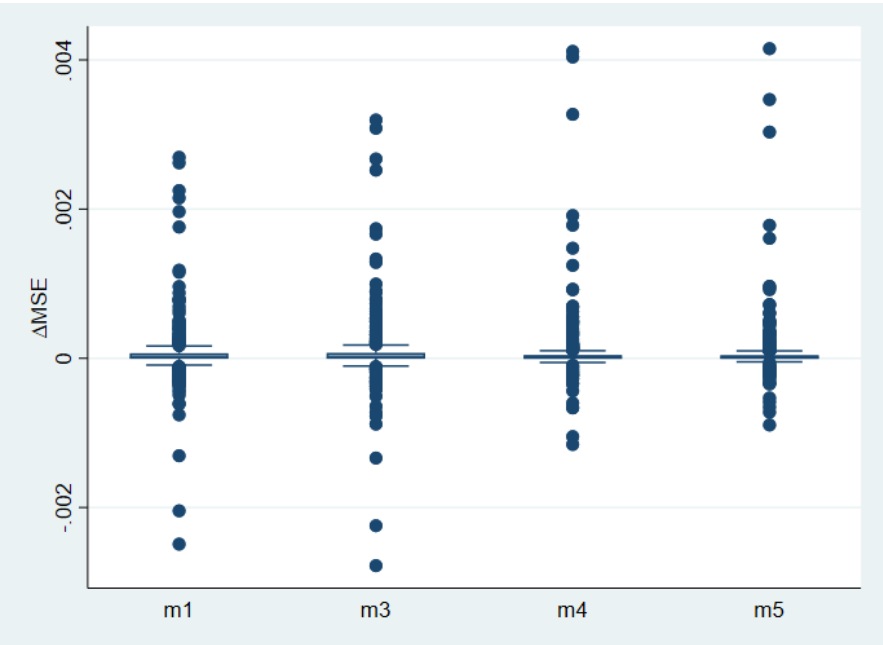

Figure 3 **Boxplot of the change in MSE for estimates of the proportion in employment and/or study for each category of subject within provider, by weighting model**

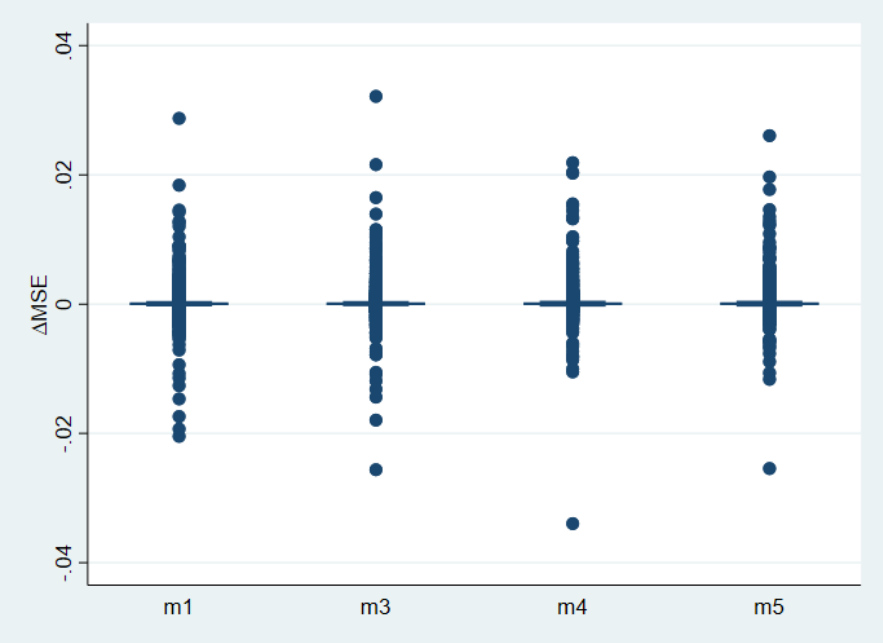

Figure 4 **Boxplot of the change in MSE for estimates of the proportion in highly skilled employment and/or study for each of 19 subject areas, by weighting model**

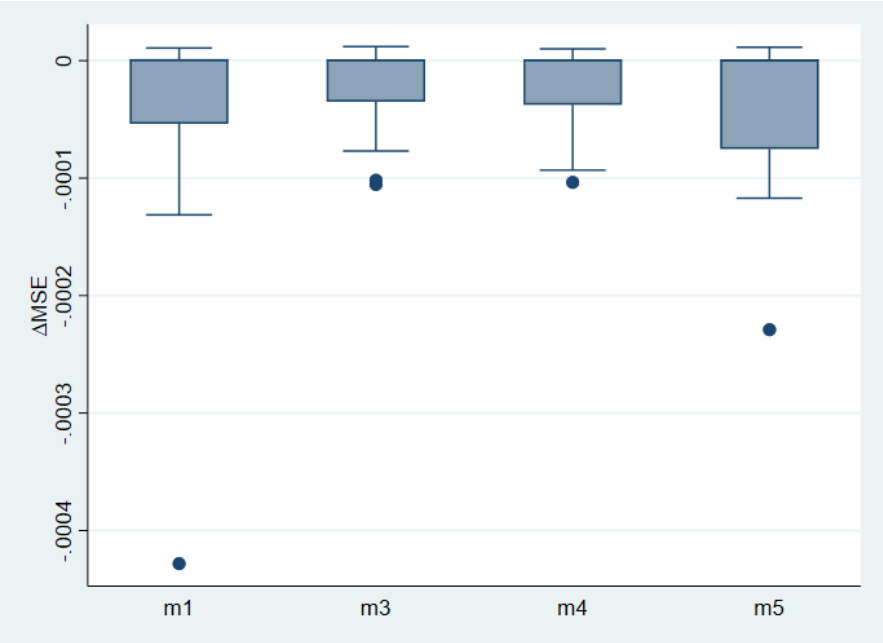

Figure 5 **Boxplot of the change in MSE for estimates of the proportion in highly skilled employment and/or study for each provider, by weighting model**

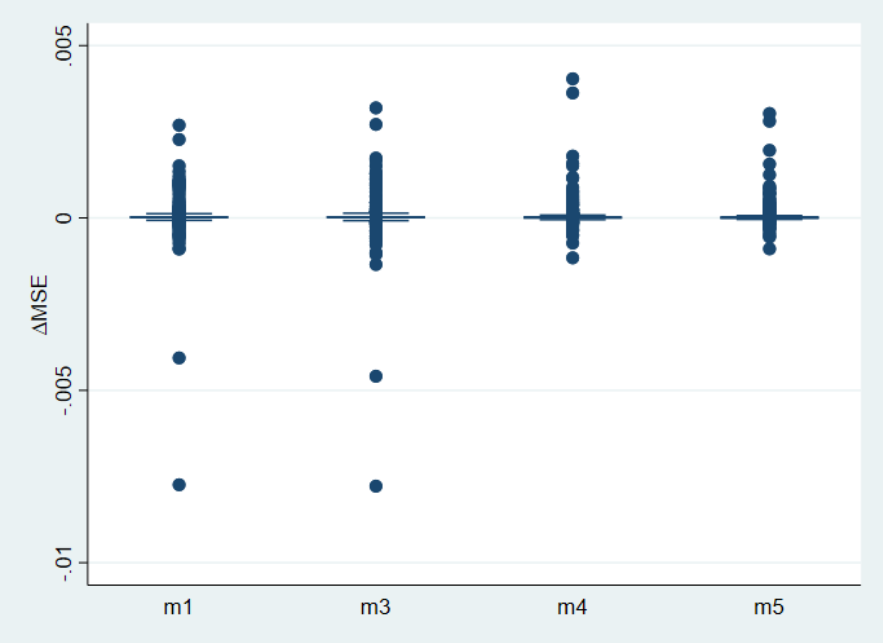

Figure 6 **Boxplot of the change in MSE for estimates of the proportion in highly skilled employment and/or study for each category of subject within provider, by weighting model**

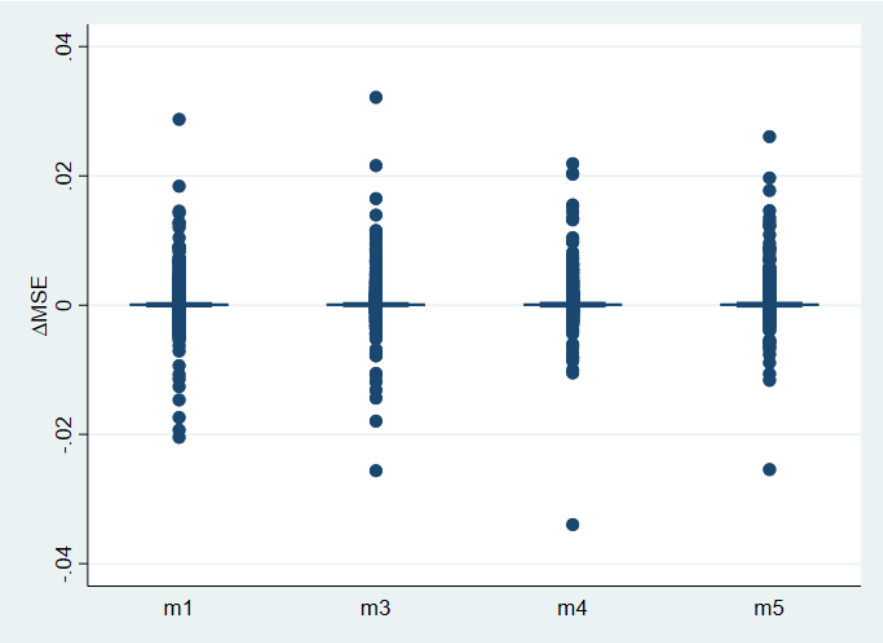

Table 8 Percentage of estimates for which weighting reduces the MSE

	Model 1	Model 3	Model 4	Model 5
y_1				
Total	22.3	21.9	14.9	16.8
UK-domiciled	24.3	23.7	14.1	14.1
Non-UK domiciled	12.0	10.2	7.4	9.1
y_2				
Total	15.7	15.2	14.7	14.9
UK-domiciled	18.9	17.5	16.4	16.9
Non-UK domiciled	10.5	9.6	6.3	6.9

Notes: y_1 in employment or study; y_2 in highly-skilled employment or study. The percentages are based on estimates for the total sample, by sex, by subject area, by provider and by subject area within provider (restricted to estimation bases with at least 100 responding graduates).

References

Biemer, P. and Christ, S. (2008) Constructing the survey weights. Chapter 16 in P.Levy and S.Lemeshow (eds.) *Sampling of Populations: Methods and Applications*. Hoboken: Wiley.

Brick, J. M. (2013). Unit nonresponse and weighting adjustments: a critical review. *Journal of Official Statistics*, 29(3):329-353.

Centre for Longitudinal Studies (2020) Millenium Cohort Study User Guide (Surveys 1 – 5) (9th edition). London: UCL Institute of Education. https://cls.ucl.ac.uk/wp-content/uploads/2020/09/MCS1-5_User_Guide_ed9_2020-08-07.pdf

Deville, J.-C., Särndal, C.-E. and Sautory, O. (2012) 'Generalized raking procedures in survey sampling', *Journal of the American Statistical Association*, Vol.88, No.423, pp. 1013-1020.

Groves, R.M. and Peytcheva, E. (2008), 'The impact of nonresponse rates on nonresponse bias: a meta analysis', *Public Opinion Quarterly,* Vol. 72, No 2, pp. 167-189. https://doi.org/10.1093/poq/nfn011

Kantar (2020) Crime Survey for England and Wales Technical report 2019/20 Volume One. London: Kantar. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeandjusticemethodology

Little, R.J.A. and Rubin, D.B. (2019), *Statistical Analysis with Missing Data*, third edition. Wiley, Hoboken NJ.

Lynn, P. (1996) 'Weighting for Survey Non-Response', 205-214, in *Survey and Statistical Computing 1996* (eds R. Banks et al), Association for Statistical Computing, Chesham.

Lynn, P. (2005) 'Weighting', 967-973, in *Encyclopedia of Social Measurement* (ed. K. Kempf-Leonard), Academic Press.

Lynn, P. (2008) 'The problem of non-response', chapter 3, 35-55, in *International Handbook of Survey Research Methods* (ed.s E.D. de Leeuw, J.J. Hox & D.A. Dillman), Routledge. https://www.routledge.com/International-Handbook-of-Survey-Methodology/Leeuw-Hox-Dillman/p/book/9780805857535

Lynn, P. (2017) 'From standardised to targeted survey procedures for tackling non-response and attrition'. *Survey Research Methods*, 11(1): 93-103. http://dx.doi.org/10.18148/srm/2017.v11i1.6734.

Nathwani, T. and Bermingham, J. (2020), *Should we Weight? Year One of the Graduate Outcomes Survey*. Higher Education Statistics Agency, Cheltenham. https://www.hesa.ac.uk/files/Graduate-Outcomes-Should-we-weight-20200521.pdf

Office for National Statistics (2020) SOC 2020 Volume 1: structure and descriptions of unit groups. Office for National Statistics, London. https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume1structureanddescriptionsofunitgroups

Schouten, B., Peytchev, A. and Wagner, J. (2017) Adaptive Survey Design. Chapman and Hall/CRC. https://doi.org/10.1201/9781315153964

Understanding Society (2020) *Understanding Society Main Survey User Guide*. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/

Appendix 1

Response rates and population distribution, 2018/19 Graduate Outcomes Survey, by sample subgroup

Subgroup		Response rate (%)	Base (population)
Total		51.6	793,185
Sex	Male	51.0	332,870
	Female	52.0	459,335
	Other	53.3	780
	Unknown	47.5	200
Level of study and class of degree	Post-graduate research	57.6	28,835
	Post-graduate taught	48.2	254,915
	Undergraduate: First class	61.7	115,065
	Undergraduate: Upper 2nd	55.4	196,325
	Undergraduate: Lower 2nd	46.9	76,740
	Undergraduate: Third class	40.6	15,460
	Undergraduate: Unclassified	54.9	19,040
	Undergraduate: Other	42.3	82,010
	Missing	48.0	4,795
Mode of study	Full-time	51.5	672,520
	Part-time	51.8	120,665
Provider type	Higher Education provider	52.1	734,755
	English Further Education College	41.9	31,865
	Northern Ireland FE College	48.0	4,795
	Alternative provider	49.1	21,775
Country of provider	England	51.4	663,660
	Northern Ireland	50.2	21,130
	Scotland	52.7	71,265
	Wales	52.6	37,130
Country of domicile	United Kingdom	56.5	579,430
	Non-UK	38.3	213,755

Subgroup		Response rate (%)	Base (population)
Subject area	Agriculture & related subjects	56.0	7,435
	Architecture, building & planning	51.3	20,040
	Biological sciences	55.9	70,735
	Business & administrative studies	42.6	140,740
	Combined	49.5	3,760
	Computer science	53.6	31,990
	Creative arts & design	50.7	68,980
	Education	53.6	64,625
	Engineering & technology	50.9	56,360
	Historical & philosophical studies	57.0	26,290
	Languages	54.1	30,035
	Law	48.1	39,995
	Mass communications and documentation	49.1	19,250
	Mathematical sciences	54.6	12,940
	Medicine & dentistry	58.0	16,510
	Physical sciences	60.5	27,165
	Social studies	52.5	78,625
	Subjects allied to medicine	55.9	76,335
	Veterinary science	59.2	1,375

Response rates and population distribution, 2018/19 Graduate Outcomes Survey, by sample subgroup (UK-domiciled graduates only) subgroup

Subgroup		Response rate (%)	Base (population)
Total		56.5	579,430
Age on entry	18 or under	60.7	170.435
	19	57.2	83,920
	20	52.9	28,220
	21 to 23	54.9	90,515
	24 to 25	53.4	32,125
	26 to 31	52.3	53,125
	32 to 35	54.8	22,605
	36 or over	61.4	62,255
Disability	Known disability	59.2	90,250
	No known disability	56.0	489,110
Ethnicity	White	56.5	443,265
	Indian	60.9	18,450
	Pakistani	54.4	16,595
	Bangladeshi	55.6	8,110
	Chinese	55.7	5,200
	Black African	58.6	27,585
	Black Caribbean	57.7	7,545
	Other	54.4	42,940
	Unknown	53.2	9,740