

Reweighting the General Household Survey 1979–2007

Éva Beaujouan^{*1}, James J. Brown^{*2} and Máire Ní Bhrolcháin^{*1}

^{*1} ESRC Centre for Population Change, ^{*2} Southampton Statistical Sciences Research Institute

University of Southampton

Abstract

We have calculated two new sets of weights applicable to the General Household Survey (GHS) from 1979 to 2007. One of these is for use with any general analysis of GHS topics and the second is designed for analyses of data collected in the Family Information section. The methods used follow closely those employed by ONS from 1996 onwards. The performance of the weights is assessed in estimating the Total Fertility Rate (TFR) from 1971-2007, an aggregate measure of fertility for which reliable figures are available at national level from vital registration statistics. Our weights improve the GHS estimates, reducing bias both in the TFR and in age-specific fertility rates.

Introduction

This article describes the construction of two consistent sets of weights for the General Household Survey (GHS) from 1979 to 2007, one for use with the whole dataset and the second for use in analysing the Family Information section, in which information on demographic histories is collected. The ESRC Centre for Population Change at the University of Southampton has constructed a time-series of 27 years of the General Household Survey, between 1979 and 2007. A central purpose of the dataset is to carry out research on demographic trends in Britain in recent decades, particularly on change over time in fertility, partnership and household composition. The CPC dataset includes, in harmonised form, substantial parts of the demographic histories collected in the Family Information section together with a selection of other variables.

The ONS has calculated weights for the GHS and includes these in the datafiles provided to the ESRC Data Archive for distribution to the academic community. As the GHS sample design is self-weighting, the purpose of the ONS weights is to adjust for non-response and to calibrate to population totals by age and sex. However, they have two shortcomings for the analysis of the Family Information (FI) section of the GHS time series:

1. They are available for the survey years 1996–2007 only, and so are lacking for 18 of the 27 GHS rounds included in the CPC time series datafile. This is problematic, as non-response, both at household level and individual level for the Family Information section, was not negligible prior to 1996¹.
2. The ONS weights are designed for analysing all persons in the sampled households, and not for the subset of eligible women and men who ultimately respond to the Family Information section and provide demographic history information.

Hence, we required weights that were on a consistent basis across the time series both for the sample as a whole and for the subsample of respondents to the Family Information section.

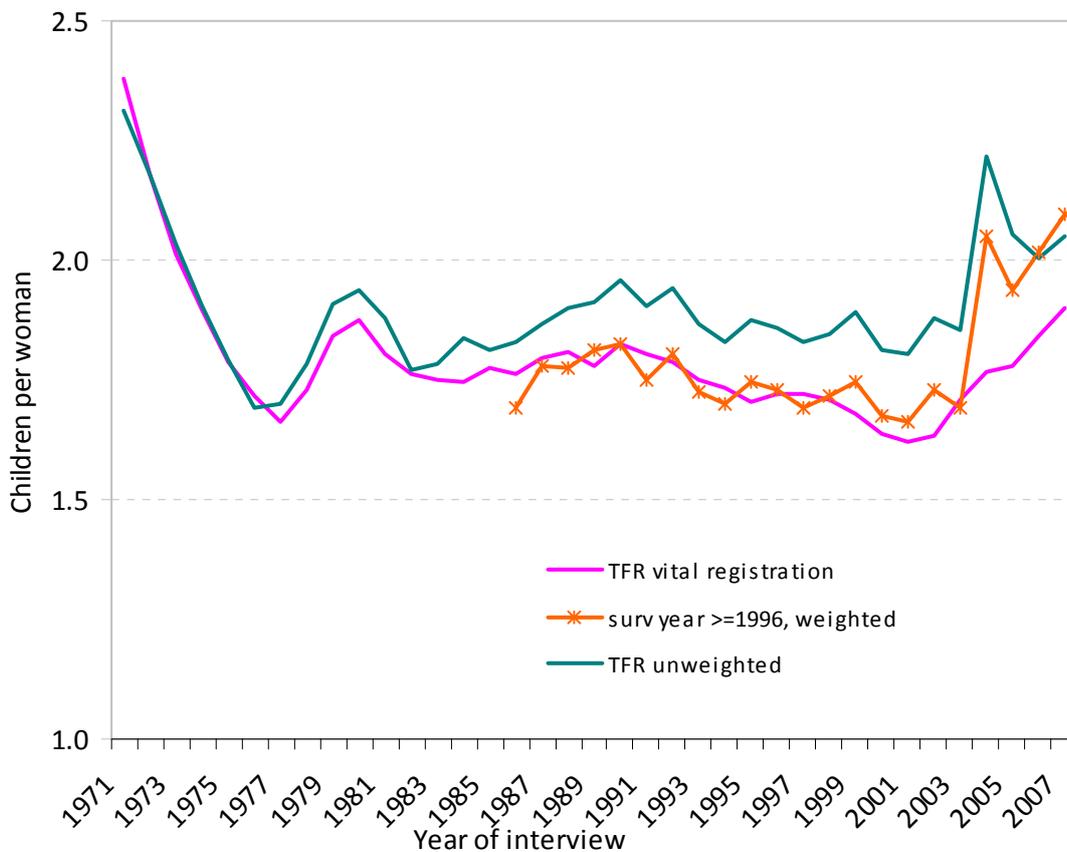
The article is organised as follows. We first review some data motivating the construction of new sets of consistent weights for the whole time-series 1979–2007. We then describe the methods used to generate weights for the household sample, and for the Family Information section subsample. Finally, we evaluate our new weights by comparing the weighted GHS estimates of a time series of the annual total fertility rate (TFR) against vital registration data.

Rationale for reweighting the CPC GHS time-series

Our primary research interest is in analysing time-trends in fertility, partnership and household change. Our research objectives include both estimating population level parameters and the conditional modelling of fertility and other demographic features as dependent variables. However, unweighted GHS estimates of national level fertility parameters evaluated against national vital registration statistics are upwardly biased, a problem that is particularly severe in recent years². This is seen in **Figure 1** which shows the estimated annual total fertility rate (TFR) of Great Britain (GB) for 1971–2007 based on the GHS fertility histories, compared with the GB TFR based on vital registration statistics^{3,4,5}. Note that the GHS estimates of the TFR are obtained by combining the GHS retrospective birth histories in all rounds specified; this means that fertility histories collected

in any given round contribute to the estimated TFR in all previous years for which FI respondents in that GHS round were exposed to childbearing. The blue line in Figure 1 shows unweighted estimates of the TFR going back to 1971, and clearly shows a positive bias from the late 1970s on. The orange line is based only on the years for which ONS weights are available, that is, 1996–2007, and plots weighted estimates from 1986 onwards. The ONS weights from the 1996 round onwards are clearly effective in reducing the bias in the estimated TFR from 1986 to 2003. The evidence of Figure 1 and the level of non-response prior to 1996 together suggested that a promising strategy for reducing the bias in the earlier part of the series would be to obtain weights for the GHS rounds prior to 1996. Furthermore, household non-response (middle response rate) prior to 1996 ranged between 24 and 33 per cent from 1996 to 2007 and from 15 to 20 per cent between 1979 and 1995¹, and so was high enough both to suggest that non-response was a source of bias and that the development of weights could potentially counteract this. As weights were to be generated anew, in order to obtain a consistent series for the entire period 1979–2007 it was necessary to reweight the whole survey series from 1979 to 2007, rather than to produce an additional set of weights for the period 1979 to 1995.

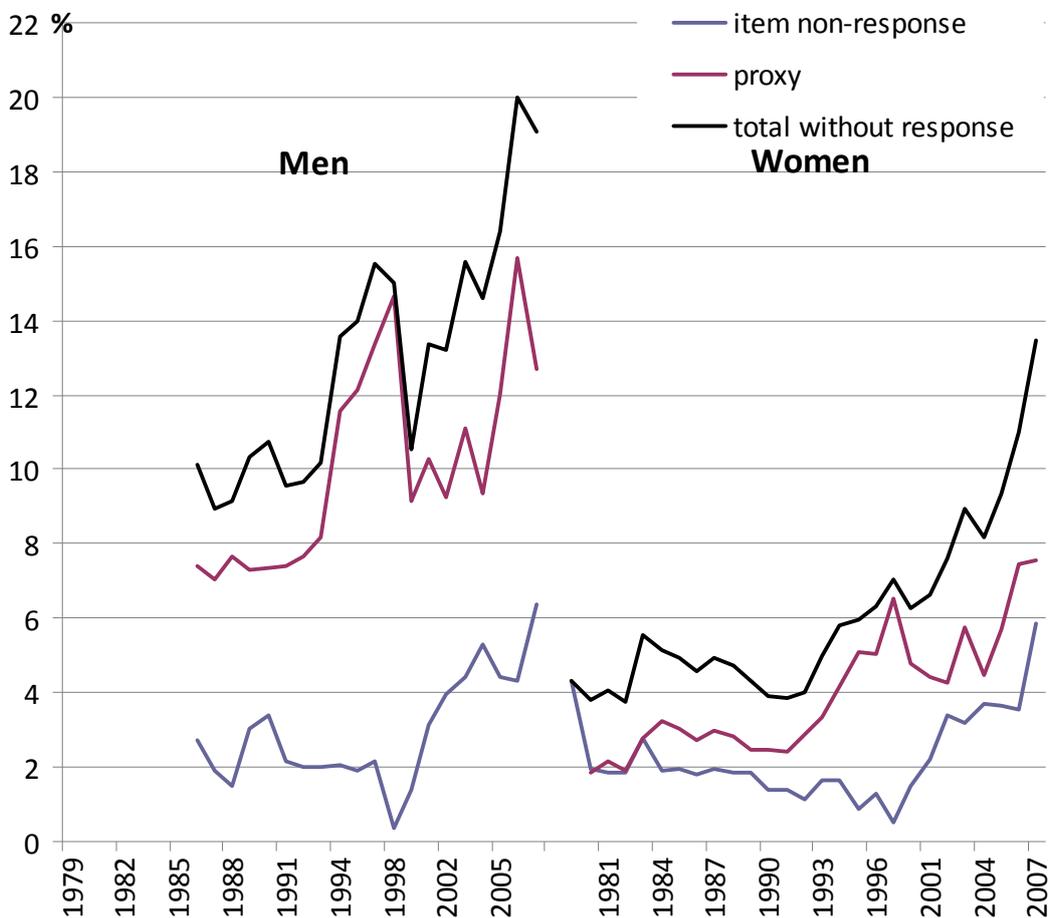
Figure 1 Weighted and unweighted GHS estimates of the total fertility rate (TFR) compared with vital registration figures, using ONS weights 1996–2007. Great Britain, 1971–2007



Note: Sample: women with a valid revised childbearing history

Beyond household level non-response, we need also to take account of individual non-response. People eligible in principle for the Family Information section are a subset of all adults in the household, with the details of eligibility varying somewhat over time. Before 1985, all women aged 18–49 and ever married women aged 16–17 were eligible for the FI section. From 1986 onwards, eligibility extended to all women aged 16–59 as well as men aged 16–59, though men were not asked about their fertility histories. Non-respondents are of two types: proxy respondents, who are not asked the FI section, and respondents who refuse the FI section outright. Also, among those answering the section, some provided such inconsistent information that they have been dropped for analytical purposes.

Figure 2 Individual level non-response to the Family Information section: per cent of proxies and of non-respondents, weighted with CPC-ALL weights. GB, GHS 1979–2007



Notes: Item non-response includes non-proxy respondents who refused the FI section outright and invalid fertility or partnership histories.

Sample: men and women eligible for the FI section, GHS 1979–2007

The level of individual non-response to the FI section is shown in **Figure 2**, separately for women and men. We see that for women, the combined level of non-response to the FI section up to the early 1990s is relatively low and usually under 5 per cent in total; it rises substantially from the mid

1990s to 2007, when the figure reaches 13.5 per cent, thus giving greater scope for individual-level bias at the end of the series. Among men, the total non-response to the FI section is much higher, around 10 per cent up to the mid 1990s, and also increasing precipitately thereafter up to 20 per cent, again underlining the potential for greater bias in later years of the series. However, the household weights do not take account of the characteristics of proxies, or of respondents to the main survey who refuse the FI section outright, or of respondents whose FI data are unusable. As some information is available from the main questionnaire on all such non-respondents, it is possible to use it to supplement the household weights and adjust further for non-respondents who are eligible in principle for the section but for whom no usable FI data exist. With this additional information, a set of weights was generated that is intended to adjust for individual non-response to the FI section, over and above the household level non-response.

In summary, there were good grounds, and data available, for constructing two consistent sets of weights for 1979–2007, one designed for the GHS sample as a whole and the second designed to adjust additionally for individual non-response to the FI section.

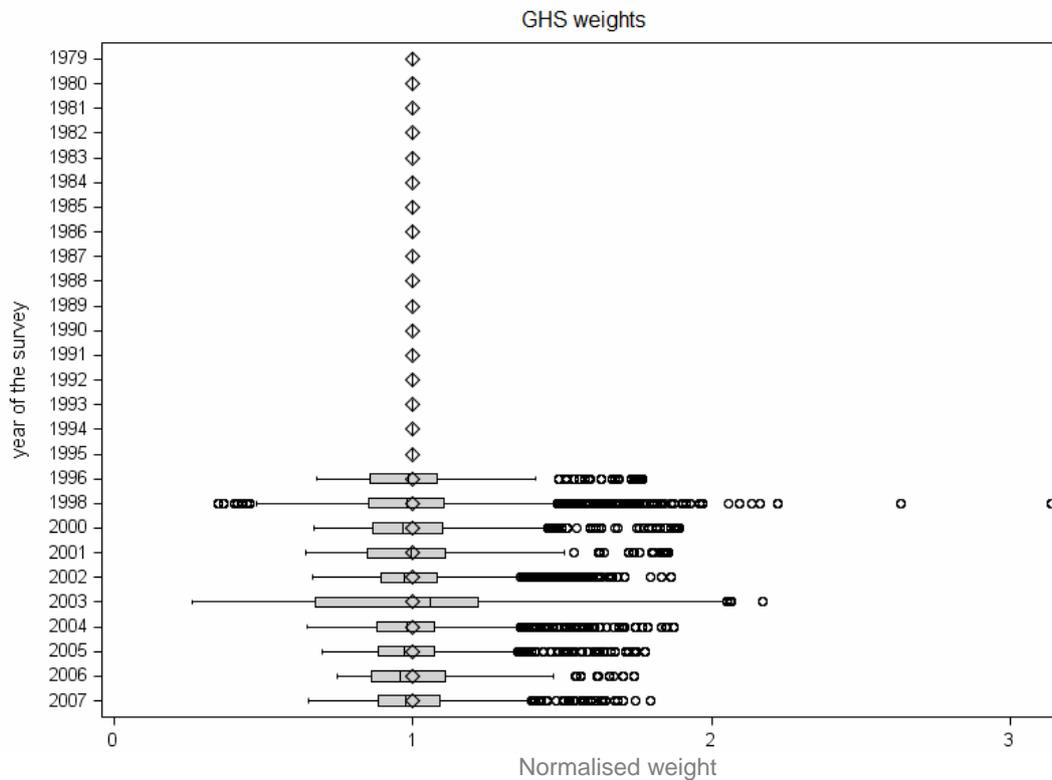
The next section gives a brief description of the method by which ONS calculated the GHS weights for 1996–2007 and some details as to their characteristics and relationship to key analytic categories.

Original ONS weights

The ONS weights are produced in two phases: firstly, non-response weights for households obtained from the GHS-census link study are applied to the GHS data, the weighting classes being identified via CHAID; secondly, the dataset is calibrated to the population distribution by age and sex, and separately by region, using Calmar from 1996 to 2002 and GES from 2003 to 2007 (for an explanation of these terms, see the **Glossary, Box 1**). Two sets of non-response factors for the GHS exist, one based on the 1991 Census-GHS link and one on the 2001 Census-GHS link. The 1991-based factors are used by ONS in weighting the 1996–2005 GHS rounds and the 2001-based factors from 2006 onwards. These weighting categories differ somewhat between the two census links and are reproduced in **Appendix Tables A1 and A2**. Further details are given in Barton⁶ and in Appendix D of Ali et al¹. The classes for calibration purposes changed in minor ways from 2004 onwards, and are shown in Appendix **Tables A3 and A4**.

The distribution for all individuals of the normalised ONS weights for each year is displayed using box plots in **Figure 3**. Each individual has a weight of 1 before 1996, as the sample design was self-weighting and no adjustment was made for non-response prior to 1996. The diamonds give the mean, the rectangle the 25th, 50th and 75th percentile, the additional lines the minimum and the maximum, and the circles the outliers. We see that the ONS normalised weights are not too dispersed with the vast majority well below a value of 2.0.

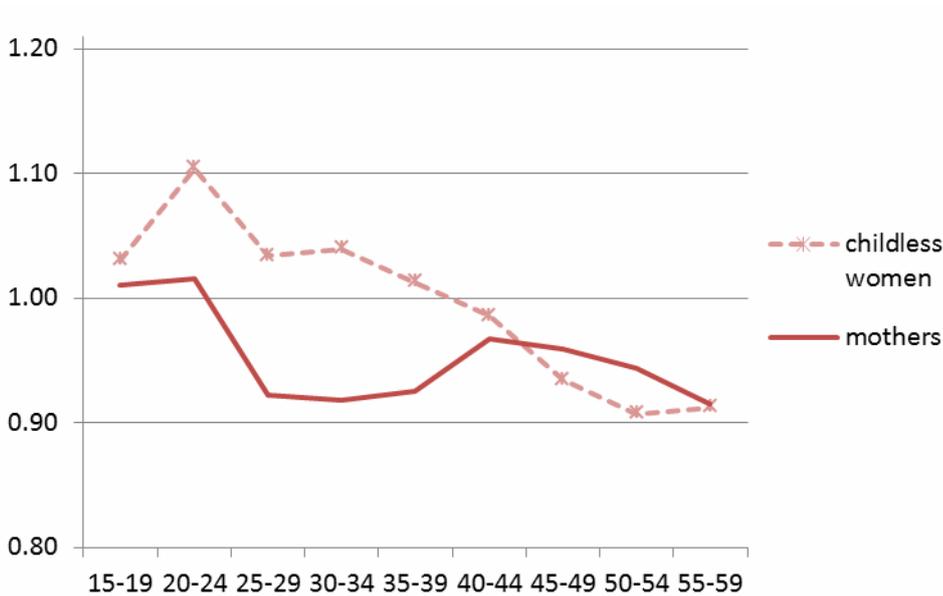
Figure 3 **Boxplot of the ONS normalised weights for the whole sample, untrimmed. GB, GHS 1996–2007**



Variation in the ONS weights with parental status and age

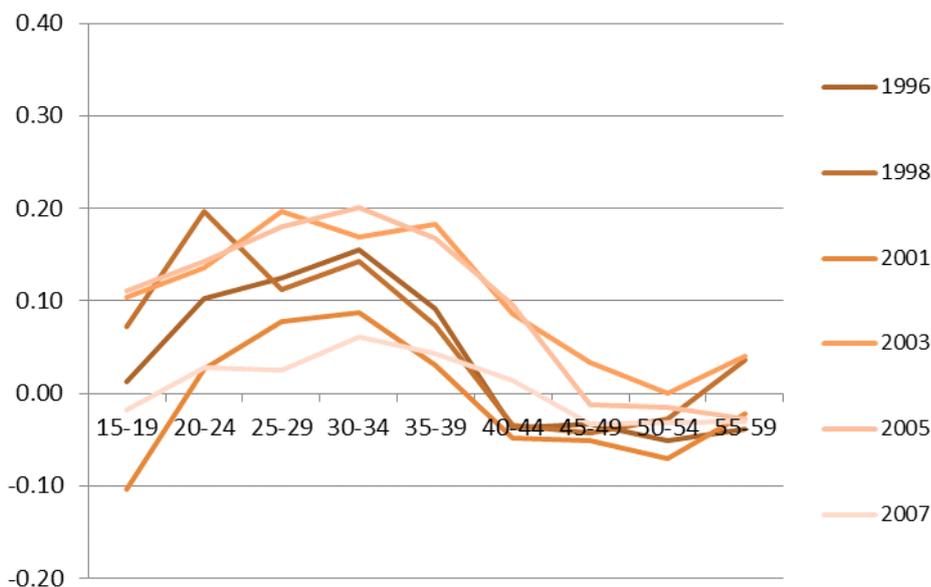
We saw above that the ONS GHS weights from 1996 onwards have a substantial effect in reducing the positive bias in the unweighted GHS estimates of the TFR. The origin of this effect is in the differential weighting of women by parental status and age. As we would expect from the higher levels of non-response among childless women, who are less likely to be at home than women with children, in all years, childless women have higher average (normalised) ONS weights than mothers between age 15 and 39 (**Figure 4a**). This contrast is found on all years, more or less strongly, as **Figure 4b** shows; it pictures the difference between the average weights of childless and parous women at each age. The differential can be sizeable - for instance, the average weighting applied to childless women aged 30–34 is 0.2 points higher than the equivalent one for mothers, which corresponds to around 15 per cent higher weighting. The differential in the weights differs substantially by age, childless women having higher weights roughly before age 45 and lower after that age. This patterning of the weights corresponds to what would be expected, with younger childless women particularly prone to survey non-response. Childless women at younger ages are, by implication, underrepresented in the GHS sample. This explains the upward bias in the TFR seen in Figure 1 above, and Figure 4 reveals the mechanism by which the weighting reduces the bias^{7,8,9}. These figures add to the rationale presented in the preceding section for seeking to reduce the bias in earlier years of the survey by generating new series of weights.

Figure 4a Average ONS weights of childless women and of mothers by age group (averages over the period 1996–2007). GHS 1996–2007



Note: Sample: women with a valid revised childbearing history

Figure 4b Difference in mean ONS weights, childless women minus mothers, by age group GHS selected years between 1996 and 2007



Note: Sample: women with a valid revised childbearing history

The combination of household non-response weights and calibration to population totals in the ONS weights alters the relative weighting of the various household types, and thus the relative

weights of households with and without children. The primary contributor to this differential is the raking procedure, rather than the application of the household non-response weights. This is because the raking is up-weighting young people, but doing so differentially in households where there are only young people, rather than in those with a mixture of younger and older. As a result, younger childless women are up-weighted. The pattern is what would be expected in order to correct for the greater response rates of women with children in the household. The overall pattern of weights is thus reasonable. However, the strong contrast between the weights of childless women and of women with children suggests that for any analysis of fertility, a consistent set of weights throughout the series covered is crucial.

Construction of CPC weights

The present section describes the construction for 1979 to 2007 of the two sets of CPC weights:

1. Weights applicable to the whole household population, suitable when using the entire sample, referred to for economy in the present article as CPC-ALL.
2. Weights applicable to the subsample eligible for and who gave valid information in the FI section, referred to as CPC-FI.

Weights for the whole sample: household non-response and adjustment

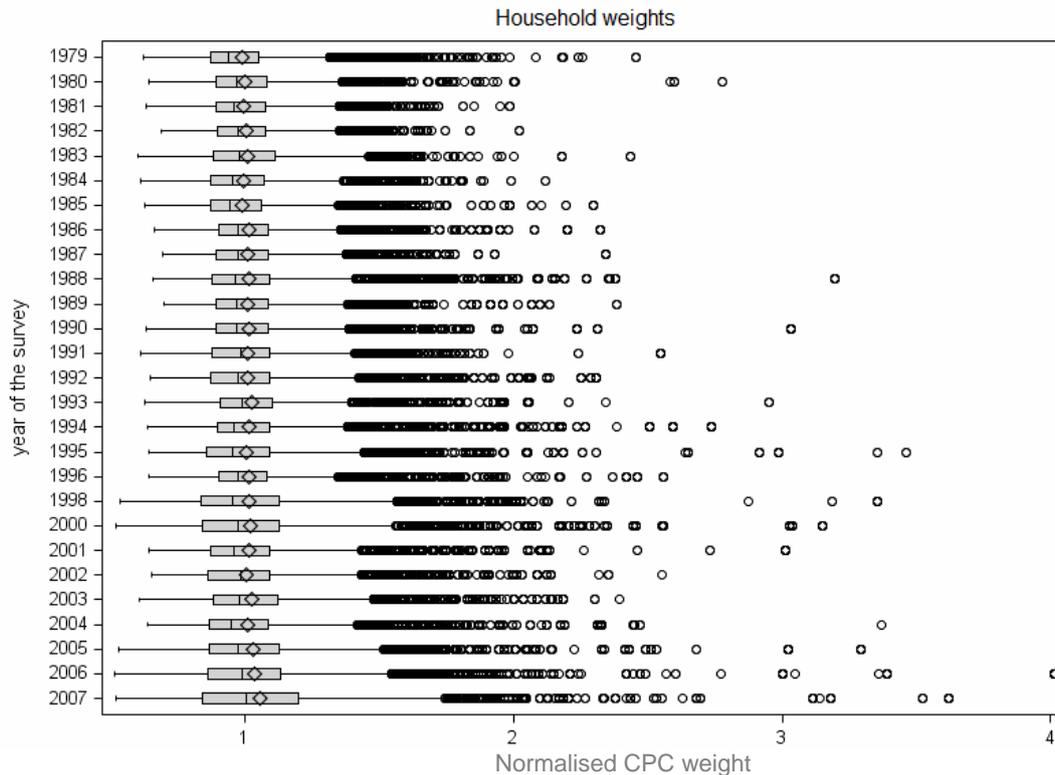
The CPC weights covering the whole survey sample are derived by broadly the same two-stage procedure used by ONS: first applying the ONS household non-response factors, and then calibrating to population totals by age and sex, and separately by region. We extend the weighting back to the 1979 survey, and also introduce the 2001-Census link non-response weights from 1998 onwards, rather than from 2006 on as in the ONS weights.

To adjust for household-level non-response we use the census-linked non-response weights already calculated by ONS for GHS, using the weighting classes given in Appendix Tables A1 and A2. As there was no GHS-Census link study in 1981, we have used the non-response weights derived from the 1991-Census link study back to 1979 and up to 1996. We have no evidence on whether the characteristics of non-responding households remained the same between 1979 and 1991. They are unlikely to have undergone major change during that period, especially as survey structure remained largely the same, and the non-response rate itself was relatively stable during those years. We use the 2001-based factors from 1998 to 2007, rather than from 2006, as ONS has done. This choice is justified by the net contrast in survey middle response rate as seen in Ali et al.¹: between 1995 and 2000, the proportion of people answering the survey dropped from 80 per cent+ to around 70 per cent. An additional reason is that before 1998 we do not have all the necessary variables in our dataset to use to the 2001-based factors, the classification by region having changed from 1998 onwards. Our choice was determined by this constraint, but appears reasonable given both the changes in non-response occurring in that period, and the break in the survey series around this time¹⁰. Overall, we would expect our procedure to improve on the ONS weights, as we use the 1991 factors only to 1996, and use the 2001 factors from 1998 onwards.

In the second stage, we calibrate each year separately using Calmar. Raking is carried out at the household level, with an individual component. Four rather than five regions are differentiated for households (London, rest of England, Wales, Scotland) as we cannot distinguish other-

metropolitan and other non-metropolitan in the datafile made available by the ESRC Data Archive. The number of persons in the household for each age-sex category brings in the individual level component. With this method, all individuals in the same household have the same weight, and the weights of those individuals reflect both household composition and individual level characteristics (Sautory, 1993).

Figure 5 Boxplot of the normalised CPC weights covering the whole sample, untrimmed. GB, GHS 1979–2007



The non-response factors are applied before raking, which is carried out using Calmar^{11,12,13,14,15}. Related to the specific question of the efficiency of using a combination of household and individual level characteristics rather than individual weights, we can refer to the study by Lemaître and Dufour¹⁶, who conclude that their preliminary results ‘suggest that individual and family estimation could be integrated at little or no loss in efficiency for estimates of persons’.

Boxplots of the weights obtained by this procedure are presented, normalised, for each year, in **Figure 5**. These weights are, for the years where comparison is meaningful, 1996–2007, somewhat more variable than the ONS weights, but not greatly so. Trimming has only a minor effect on the point estimates of the TFR, our criterion fertility indicator; the household level CPC-ALL weights are trimmed at a value of 2.0 and used in normalized form in all subsequent analysis. This helps to reduce variance and avoids the influence of weights with extreme values on other variables not evaluated here.

Weights for analysing the Family Information section

The household weights (CPC-ALL) described in the previous section are designed for use with the whole sample, but are less suited to analysis of specific subsamples of the GHS. As noted above, those who should respond to the FI section are a subset conditional on eligibility by age and sex, but because of individual non-response within the household, information is unavailable for proxy respondents, those who refused the section outright, and those with unusable demographic histories. The absence of these non-respondents from the analysis of FI data is not taken into account in the CPC-ALL weights, and so we have developed a further set of weights to correct for the non-response specific to the FI section, over and above that adjusted for via the CPC-ALL weights.

The CPC-FI weights have three components: first, the non-response adjustment and raking was applied at the household level; second, individual level probabilities of response to the FI section were estimated; third, a calibration to population totals for the subsample eligible for the FI section was carried out.

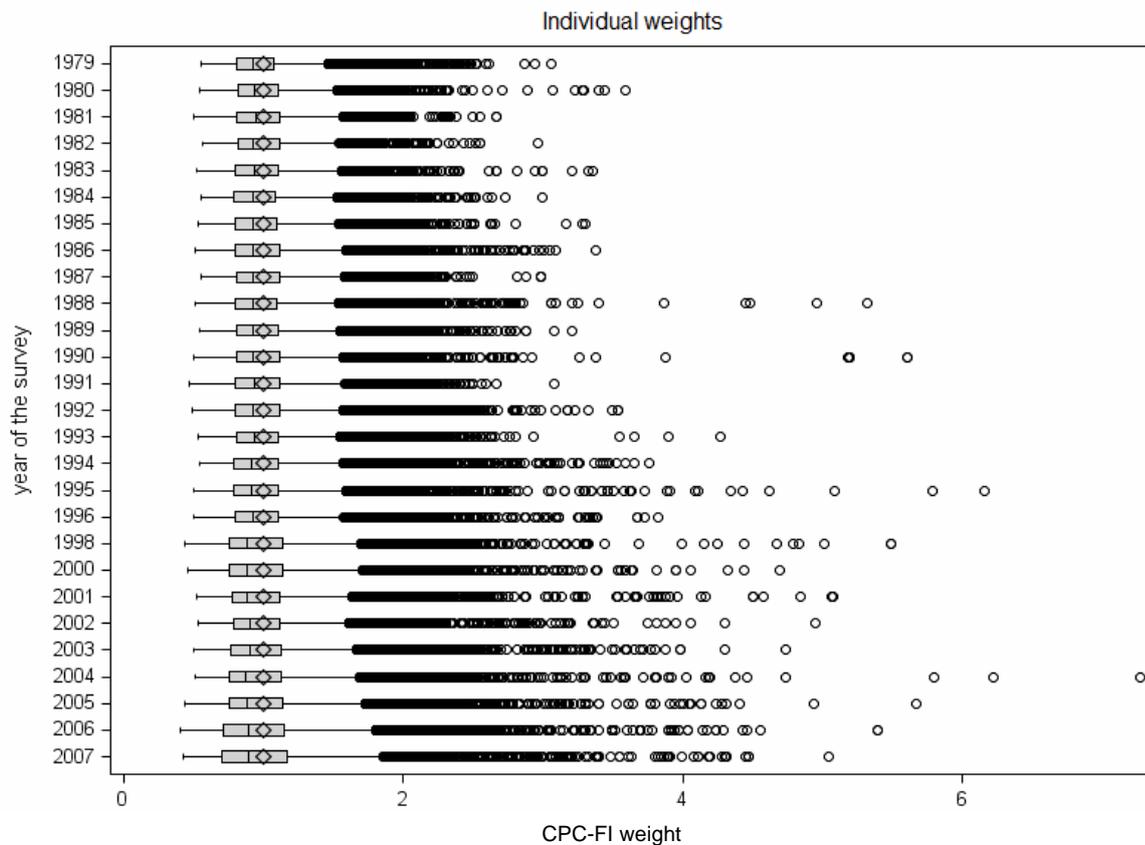
The objective of the second stage was to obtain adjustment factors for individual non-response to the FI section within households, in addition to household-level non-response. Non-response to the FI section includes, as noted above, three categories of people: proxy respondents, respondents who refused the FI section outright, and those who responded to the FI but whose demographic histories were unusable due to inconsistency or incompleteness. We have substantial amounts of information on people in responding households who did not respond to the FI section, since most of them answered most of the main GHS questionnaire, or had this answered on their behalf. Respondents are classified as having unusable FI data if their birth history (women) or their partnership history (men) was invalid.

Individual non-response weights were obtained using the response propensity scores method: a logistic regression was fitted to the probability of response and the inverse of the fitted probabilities used as weights^{17,18,19}. The predictors used in the model of response to the FI section were year, age at survey, partnership status (single, married, separated, cohabiting for the years 1986–96), number of children in the household, number of adults in the household, country of birth, age of the youngest child in the household, and age of the oldest person in the household. Several other predictors, such as tenure type and educational level, were examined but were dropped as they did not improve model fit. The response model was estimated separately for four groups of years - 1979–85, 1986–92, 1993–96 and 1998–2007 - based on the relative homogeneity of the overall level of non-response within these time periods (see Figure 2 above). Male eligibility for the FI section began from 1986, and so 1979–1985 is the first period chosen. Proxies rise substantially in 1993–1996 even though non-response decreases at the same time, though to a lesser extent. Thus we treat 1986–1992 and 1993–96 separately. Non-response increases substantially in 1998, giving the final grouping of survey years.

Importantly, the models of individual non-response to the FI section were estimated on a sample weighted by the CPC-ALL weights, and so the CPC-FI weights are based on a combination of the CPC-ALL (trimmed) weights as well as adjusting for individual level non-response to the FI section, prior to the final calibration. Furthermore, unlike the calibration for the CPC-ALL weights, the calibration to population totals for the CPC-FI weights was carried out purely at the individual level.

The normalized CPC-FI weights that result from this series of steps are shown in **Figure 6**. As would be expected from the fact that these weights now incorporate adjustment for non-response at the individual level, they are more dispersed, and extend to more extreme values than do the CPC-ALL set of weights. The effect of trimming was examined, and it was found that the weights can be trimmed to a value of 3.0 with little impact on the point estimates of the TFR.

Figure 6 Boxplot of the CPC-FI weights, specific for respondents to the Family Information section, untrimmed. GB, GHS 1979–2007

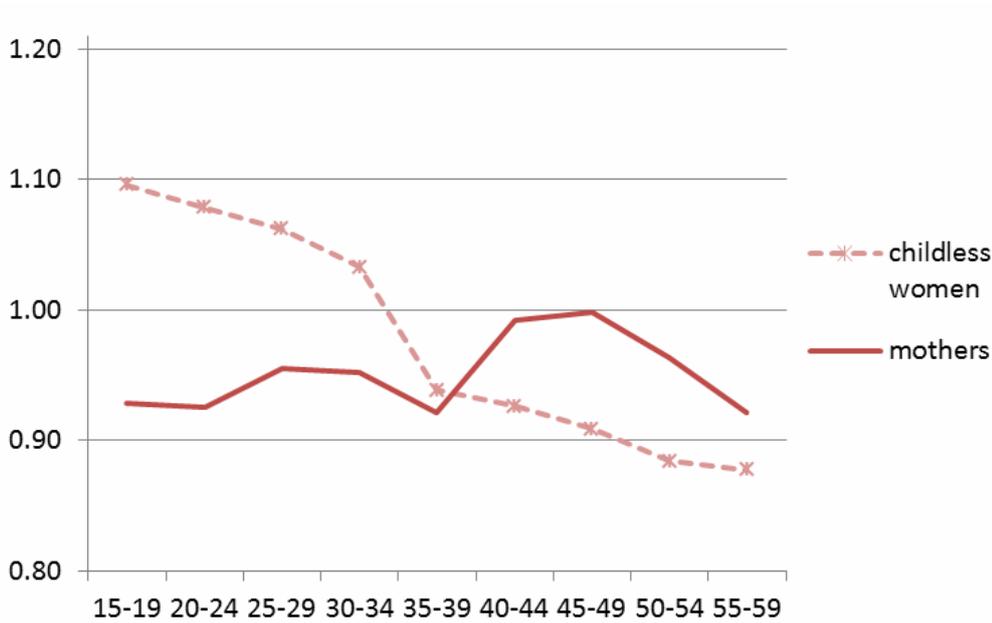


Evaluating the CPC weights for fertility estimation

Profile of weights by age and parental status

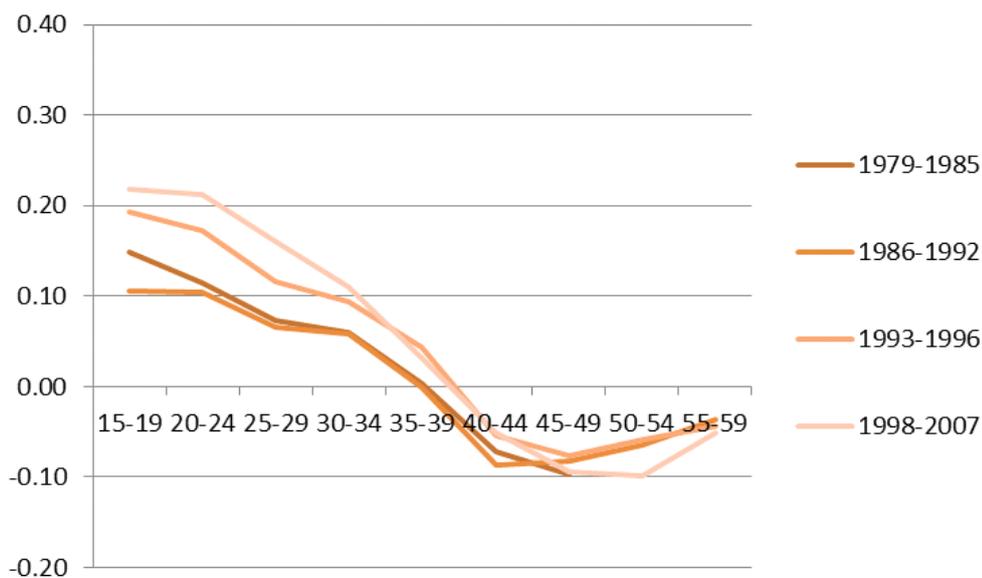
As we saw above, the GHS-weights for childless women are higher than those for mothers, at younger ages but especially in the twenties and thirties, and somewhat below at older ages. The CPC household level weights series also show a differential by parental status and age (**Figure 7a**), and this is still true when decomposing in groups of years as defined above (**Figure 7b**).

Figure 7a Average CPC-ALL weights of childless women and of mothers by age group (averages over the period 1979–2007). GB, GHS 1979–2007



Note: Sample: women with a valid revised childbearing history

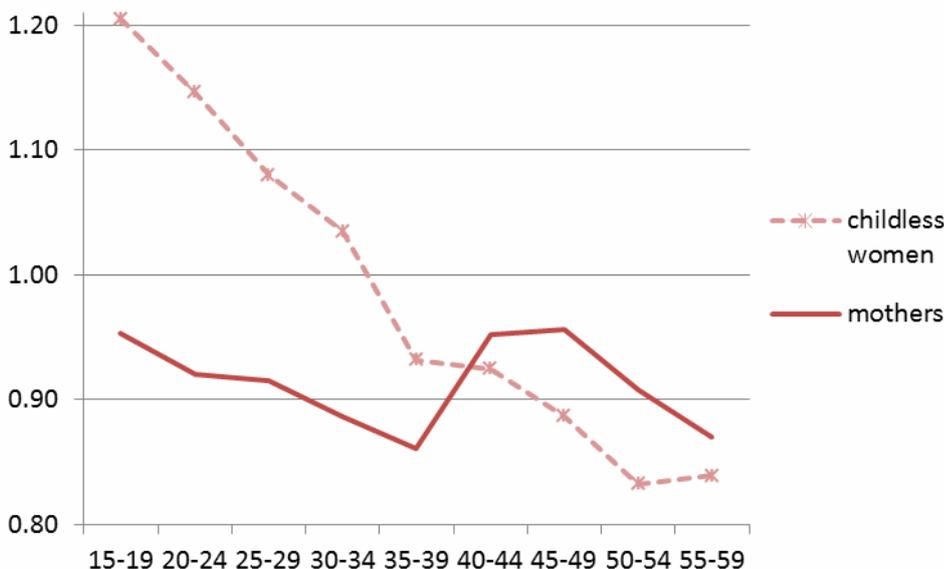
Figure 7b Difference in mean CPC-ALL weights, childless women minus mothers, by age group, for four periods. GB, GHS 1979–2007



Note: Sample: women with a valid revised childbearing history

Several features of the new CPC weights are noteworthy. First, given that we are particularly interested in fertility analysis, we showed above in Figure 4 the difference between the weights for childless women and those with children, specific by age, in the ONS sets of weights. The difference is at a peak among women in their early 30s in the ONS series of Figure 4. However, in the CPC series shown in Figure 7, the difference peaks at ages under 20. This is precisely what we would expect in view of the high non-response rate of young women, most of whom are childless. The CPC-ALL weights are, on this criterion, a decided improvement, in that the profile of differential weights by age and parenthood status is more in accord with the known high frequency of non-response among young childless women. While the differentials by age between childless women and mothers in the CPC-ALL set of weights appear more satisfactory than in the original ONS series in Figure 4, the origin of the disparity is not altogether clear. Tests indicate that it is not due to the use of the 2001 adjustment factors from an earlier GHS round in the CPC series than in the ONS series of weights. The remaining possibilities are either that the ONS weights were calibrated to population estimates that were rolled forward from the 1991 Census, while the CPC population totals are based on post-2001 Census estimates, as of early 2011, or that there is some difference between the methods as applied that we have not been able to identify.

Figure 8a **Average CPC-FI weights of childless women and of mothers by age group (averages over the period 1979–2007). GB, GHS 1979–2007**



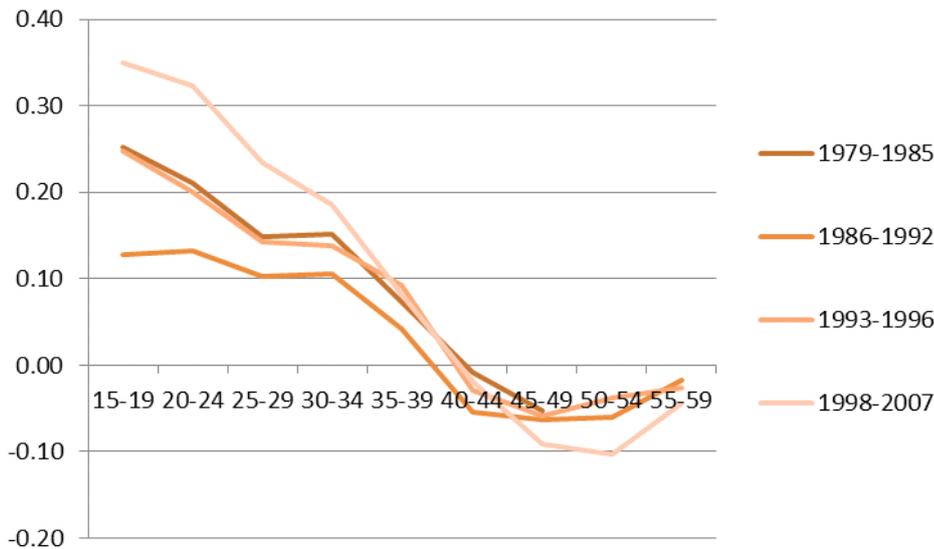
Note: Sample: women with a valid revised childbearing history

Second, the contrast between childless women and mothers becomes more pronounced at younger ages in the most recent period (compare Figure 7 with Figure 4), which is consistent with the deterioration in response rates in recent years²⁰.

Finally, both of these contrasts are even more pronounced in the CPC-FI set of weights (**Figure 8**) than in the CPC-ALL weights. This is primarily because more of the proxies and non-respondents have characteristics of childless persons, and, in particular, are disproportionately single. Almost the entirety of the change in CPC-FI weights by comparison with the CPC-ALL set is due to the

additional individual non-response adjustment, as the CPC-FI weights are modified only marginally by the raking to individual level population totals.

Figure 8b Difference in mean CPC-FI weights, childless women minus mothers, by age group, for four periods. GB, GHS 1979–2007



Note: Sample: women with a valid revised childbearing history

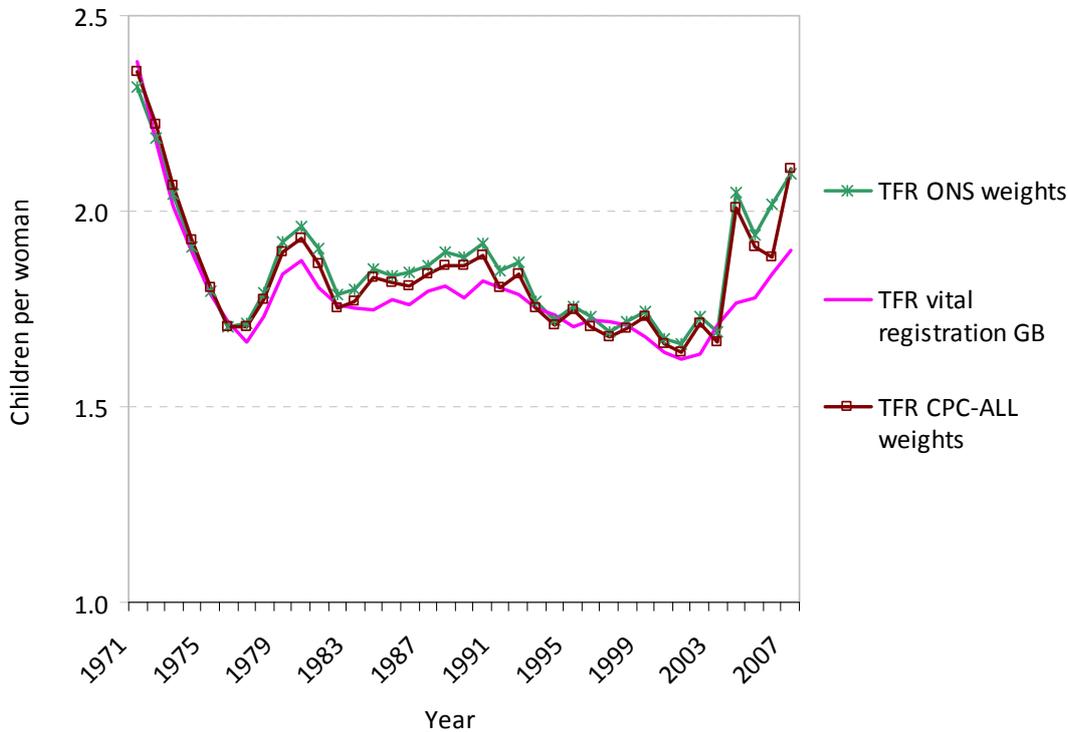
Estimates of the total fertility rate (TFR)

The GHS TFR time-series is based on birth histories of women with a valid childbearing history (aged less than 50 before 1986 and less than 60 from 1986 on). GHS rounds are pooled, and exposure to childbearing and event occurrence by age in each past calendar year are calculated using the dates of birth of women and of live births. Annual age-specific fertility rates are obtained by dividing the total weighted events at each age by the total weighted exposure at each age, in each calendar year. The annual TFR is then obtained by summing the rates at ages 15 to 44 in each calendar year.

The diagrams that follow show estimates of the TFR, comparing the results of weighting by successive pairs of weight sets, together with values from vital registration statistics.

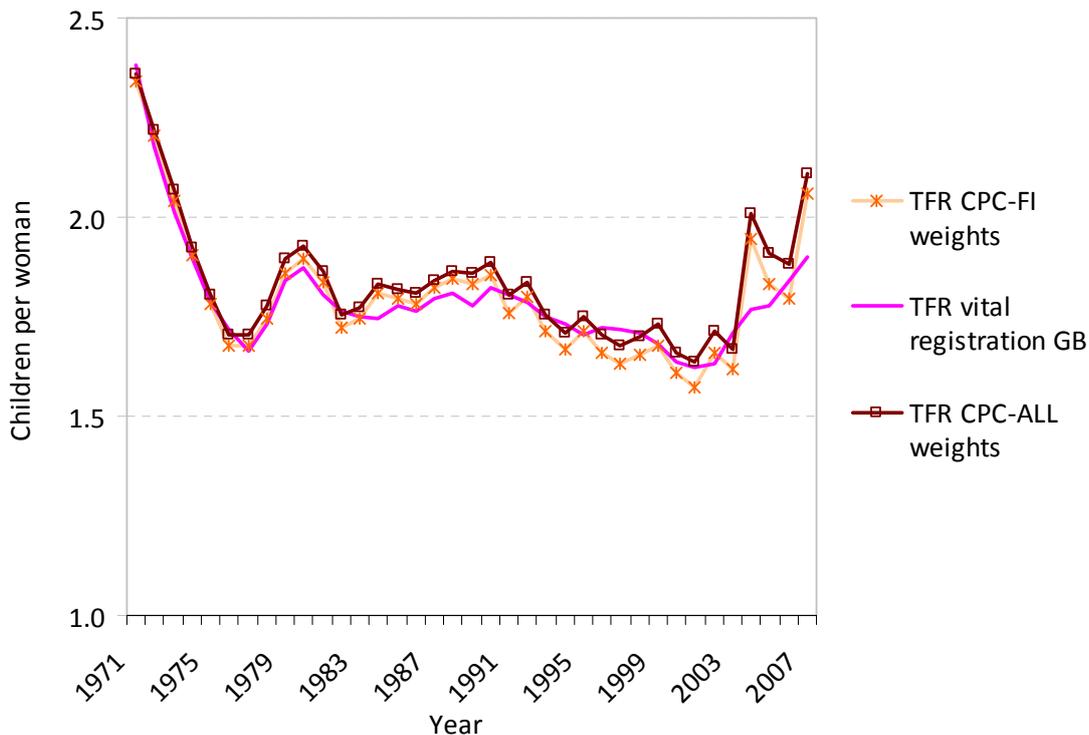
Figure 9 compares the ONS weighted series (unweighted to 1995, weighted from 1996) with the CPC-ALL weighted series and with vital registration. We see that the CPC-ALL weights reduce the bias in the ONS weighted series somewhat, though not very substantially.

Figure 9 Total fertility rate weighted with ONS weights, with CPC-ALL weights, and TFR from vital registration. GB, GHS 1979–2007



Note: Sample: women with a valid revised childbearing history; TFR calculated on ages 15-44

Figure 10 Total fertility rate weighted with CPC-ALL weights, CPC-FI weights, and TFR from vital registration. GB, GHS 1979–2007

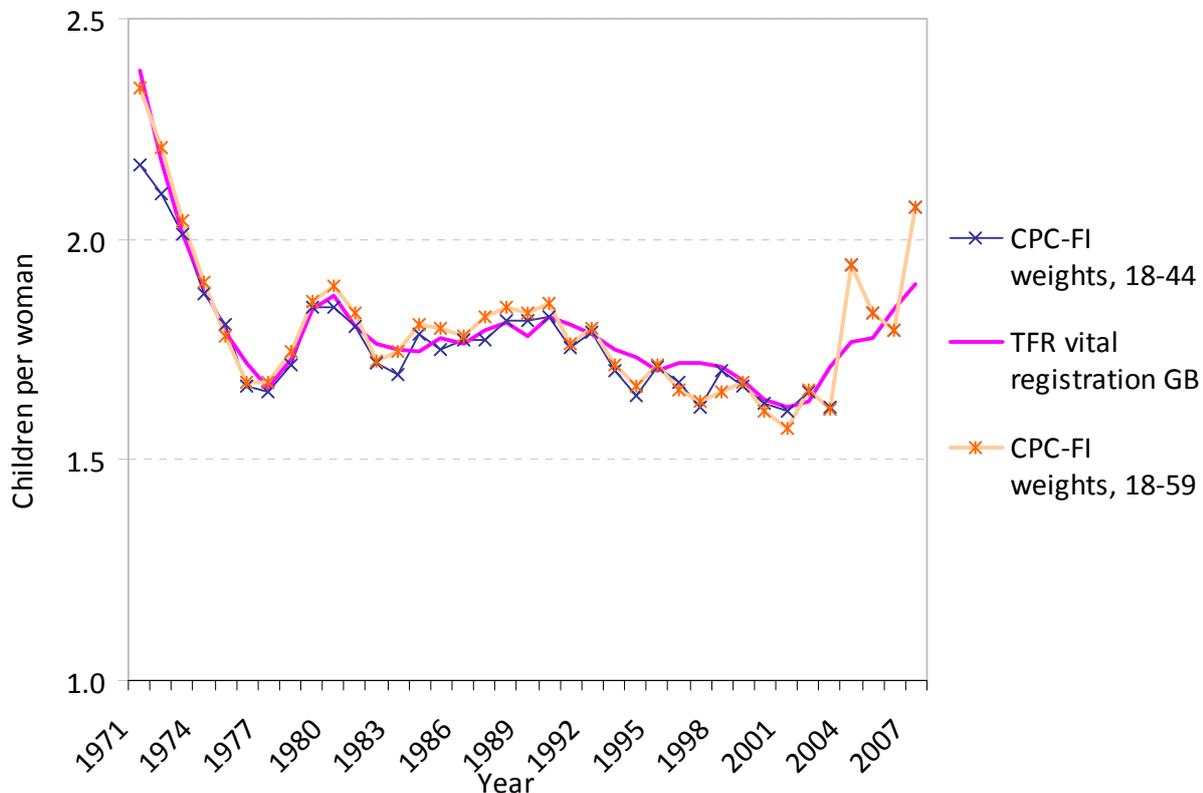


Note: Sample: women with a valid revised childbearing history; TFR calculated on ages 15-44

Figure 10 shows the series weighted by CPC-ALL and by CPC-FI weights as well as vital registration figures. The CPC-FI series improves on the CPC-ALL series in reducing the bias further.

Figure 11 shows two series weighted by CPC-FI, one of which includes all women responding to the FI section (those aged 18–59), and the other confined to women aged 18–44. Restricting the age range to 18–44 reduces the positive bias in the 1980s further. Note, however, that both weighted sets are slightly below the vital registration figure in the 1990s. These two points require some comment. The restriction to 18–44 reduces the positive bias in the 1980s because of a combination of two factors: (a) older women, particularly those aged 45+, in recent years over-reported themselves as childless (see notes 3-5) and (b) the weights for older childless women are somewhat below those of older women with children (see Figure 8 above). The inclusion of women over 45 thus creates a positive bias in the TFR for the (more distant) periods when they were in their prime childbearing years, and their exclusion reduces that positive bias. However, the slight undershooting of the TFR during the 1990s and early 2000s is, we believe, a true reflection of the content of the GHS fertility histories in more recent years, since our correction of the fertility histories using household schedule data was not quite complete⁵. Thus the fertility of women in their late 30s and early 40s at GHS rounds since 2000 is slightly underestimated, even in our revised histories. This would result in a slight underestimate of period rates in the 1990s or thereabouts, which is what we find on applying the CPC-FI weights to both the whole sample and the sub-sample aged under 45.

Figure 11 Total fertility rate weighted with the CPC-FI weights, calculated on two samples, women aged <60 and women aged <45. GB, GHS 1979–2007



Note: Sample: women with a valid revised childbearing history; TFR calculated on ages 15-44

Finally, we summarise these comparisons in **Table 1** using two measures of the error in the GHS estimate of the TFR relative to vital registration for the period 1971 to 2007. We see that the application of successive sets of weights up to CPC-FI reduces both the bias and the absolute error. Restricting the sample to the age-group under 45 introduces a small bias which is, however, expected because of the slight shortfall in births in the corrected GHS histories discussed above. For completeness, vital registration estimates of the TFR are given in **Table 2**, together with the ratios of each TFR series, unweighted and with the various sets of weights, to the vital registration figure.

Table 1 Average error in GHS estimates of annual TFR values for 1971 to 2007 compared with vital registration, unweighted and with different sets of weights. GB, GHS 1979–2007

	Mean error	Mean absolute error
Unweighted*	0.11	0.11
ONS weights*	0.06	0.07
CPC-ALL weights*	0.04	0.05
CPC-FI weights*	0.00	0.04
CPC-FI weights, 18-44**	-0.01	0.04

Notes: *Sample: women aged 18-59 at survey

**Sample: women aged 18-44 at survey

To sum up, the TFR for the period 1971 to 2007 estimated from the GHS 1979–2007 histories using the CPC-ALL weights is closer to the national figure than either the unweighted estimates or those using the ONS weights. The TFR series estimated with the CPC-FI weights reduces the positive bias in the 1980s if confined to women under 45, and slightly underestimates the vital registration figure from about 1990 to 2003. The latter feature is more consistent with the known absence of some births from the revised birth histories of recent rounds. Combined with the greater accuracy of the figures for the 1980s, we believe that the estimates based on CPC-FI weights for women aged under 45 are the best estimates currently available. These weights and their associated age range are thus the combination of choice in analysing the GHS fertility histories of the CPC time series datafile. Restricting the sample in this way has the disadvantage, however, that sample numbers are reduced and also that the length of the time series for which estimates from retrospective histories can be made is reduced also.

Table 2 **TFR from vital registration, and ratio of GHS estimates with different weights to vital registration figures. GB 1971–2007**

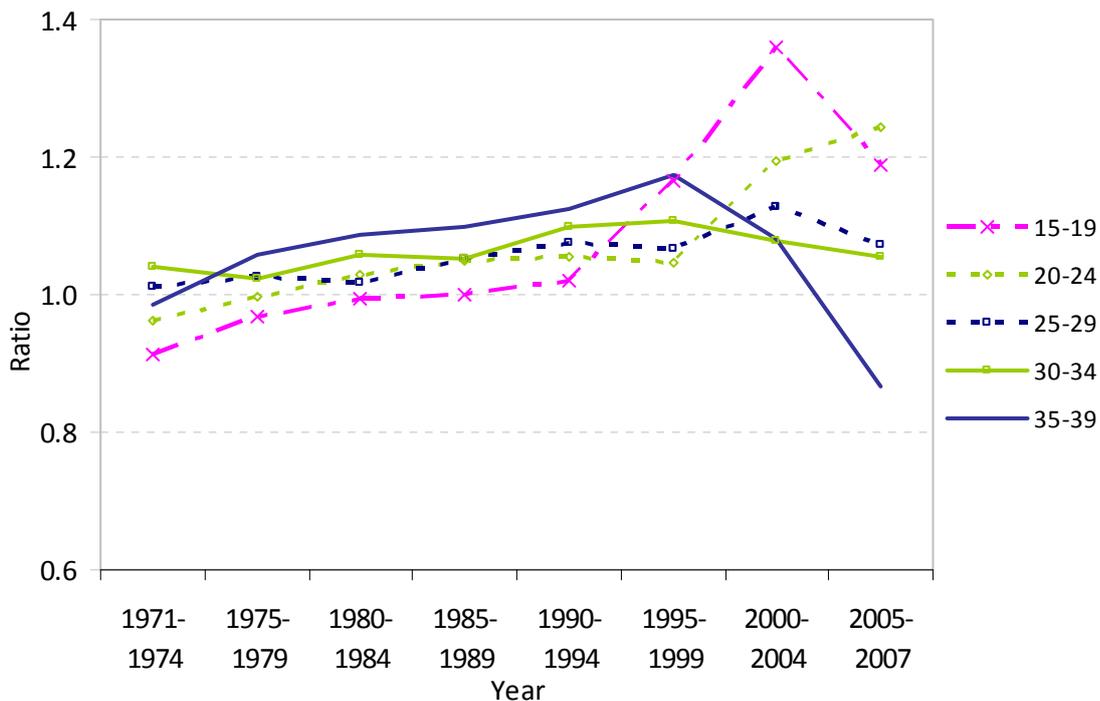
Year	TFR vital registration GB	Ratio to vital registration TFR of...				
		Unweighted	ONS weights	CPC-ALL weights	CPC-FI weights	CPC-FI weights, 18-44
1971	2.38	0.97	0.97	0.99	0.98	0.91
1972	2.18	1.00	1.00	1.02	1.01	0.97
1973	2.01	1.01	1.02	1.03	1.01	1.00
1974	1.90	1.00	1.01	1.01	1.00	0.99
1975	1.79	1.00	1.01	1.01	1.00	1.01
1976	1.72	0.98	0.99	0.99	0.98	0.97
1977	1.66	1.02	1.03	1.03	1.01	0.99
1978	1.73	1.03	1.04	1.03	1.01	0.99
1979	1.84	1.04	1.04	1.03	1.01	1.00
1980	1.87	1.04	1.05	1.03	1.01	0.99
1981	1.80	1.04	1.06	1.03	1.02	1.00
1982	1.76	1.01	1.02	1.00	0.98	0.98
1983	1.75	1.02	1.03	1.01	1.00	0.97
1984	1.75	1.05	1.06	1.05	1.03	1.02
1985	1.78	1.02	1.03	1.02	1.01	0.98
1986	1.76	1.04	1.05	1.03	1.01	1.01
1987	1.79	1.04	1.04	1.03	1.02	0.99
1988	1.81	1.05	1.05	1.03	1.02	1.00
1989	1.78	1.08	1.06	1.04	1.03	1.02
1990	1.82	1.07	1.05	1.03	1.02	1.00
1991	1.81	1.05	1.02	1.00	0.98	0.97
1992	1.79	1.09	1.05	1.03	1.01	1.00
1993	1.75	1.07	1.01	1.00	0.98	0.97
1994	1.73	1.06	0.99	0.99	0.96	0.95
1995	1.70	1.10	1.03	1.03	1.01	1.00
1996	1.72	1.08	1.00	0.99	0.96	0.97
1997	1.72	1.06	0.98	0.98	0.95	0.94
1998	1.71	1.08	1.00	0.99	0.97	0.99
1999	1.68	1.13	1.04	1.03	1.00	0.99
2000	1.64	1.11	1.02	1.01	0.98	0.99
2001	1.62	1.11	1.02	1.01	0.97	0.99
2002	1.63	1.15	1.06	1.05	1.01	1.01
2003	1.71	1.09	0.99	0.97	0.94	0.95
2004	1.77	1.25	1.16	1.14	1.10	1.10
2005	1.78	1.16	1.09	1.07	1.03	1.03
2006	1.84	1.09	1.10	1.02	0.97	0.97
2007	1.90	1.08	1.10	1.11	1.09	1.09

Age specific fertility rates

Finally, we look briefly at the impact of the weights on the age-specific fertility rates, which are the constituents of the TFR. The CPC FI-weights improve the fit between the survey based age specific fertility rates (ASFRs) and the corresponding figures from vital registration. **Figure 12** plots the ratio of the ASFRs calculated from the unweighted retrospective survey histories to the

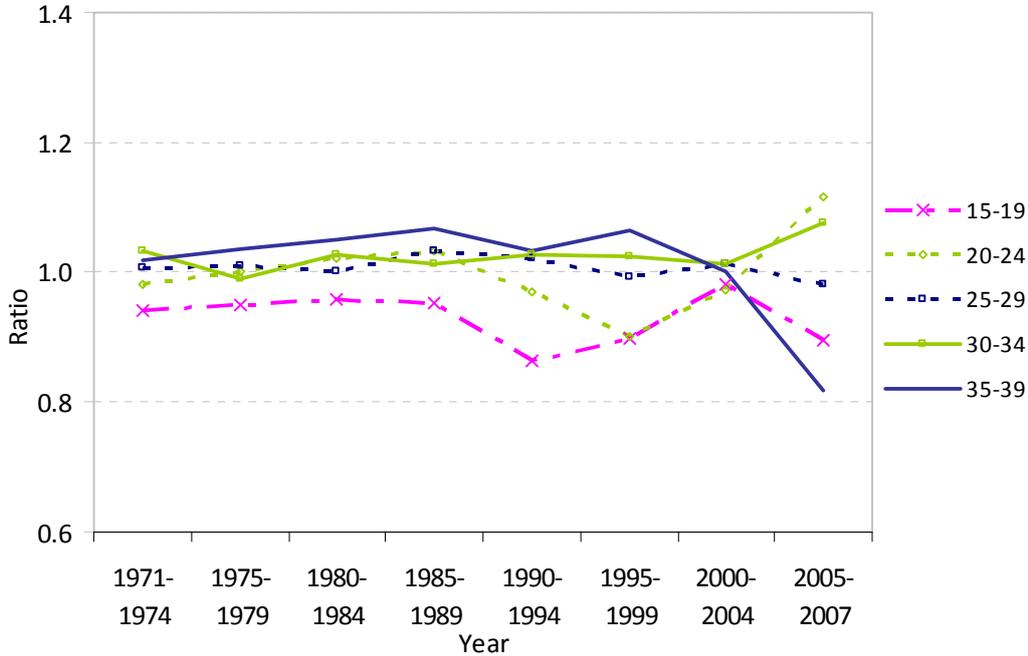
published vital registration values for England and Wales. A ratio above 1.0 reflects overestimation in the survey series relative to the national figure. As we see, the ratios are well above 1.0, and rising over that 35-year period. However, the application of the CPC-FI weights (**Figure 13**) reduces the bias in all age groups except 15–19. The unweighted GHS estimates of teenage birth rates are initially below, then close to, national levels, and finally well above vital registration figures. Weighting, however, brings the estimates well below national levels throughout, thus reflecting under-reporting of teenage births in the GHS, which is, however, offset in the unweighted estimates by the high non-response rates of this group. Years from 1980 onwards, where unweighted rates are decidedly above national levels, improve particularly on weighting, but fertility seems underestimated by our weighted time series at ages 20–24 in 1995–8, and accuracy deteriorates in the most recent rounds covered, 2005–07. Sample sizes are smaller in 2005-07 both because estimates are based only on the fertility histories of respondents in 2005-07, and because data are used here only from those respondents in 2006-07, a quarter of the total, who were interviewed for the first time (three quarters were re-interviews due to the introduction of a longitudinal structure in 2005). Finally, we show in **Figure 14** that the use of the CPC FI-weights reduces the variability of the difference between the time series and the vital registration age specific birth rates as evaluated by the root mean square error, thus providing evidence of the improvement in the quality of the fertility estimates achieved by the reweighting exercise. As any error between a survey-based estimate and the published TFR will be due to both bias and sampling variability it is encouraging to see that both the mean error and mean absolute error are reduced as it implies that the more variable CPC weights are not resulting in considerably more variable estimates but give estimates with a consistently smaller error.

Figure 12 Ratio of age specific fertility rates based on revised fertility histories to national vital registration, survey years 1979–2007, unweighted. GHS, England and Wales, 1971–4 to 2005–7



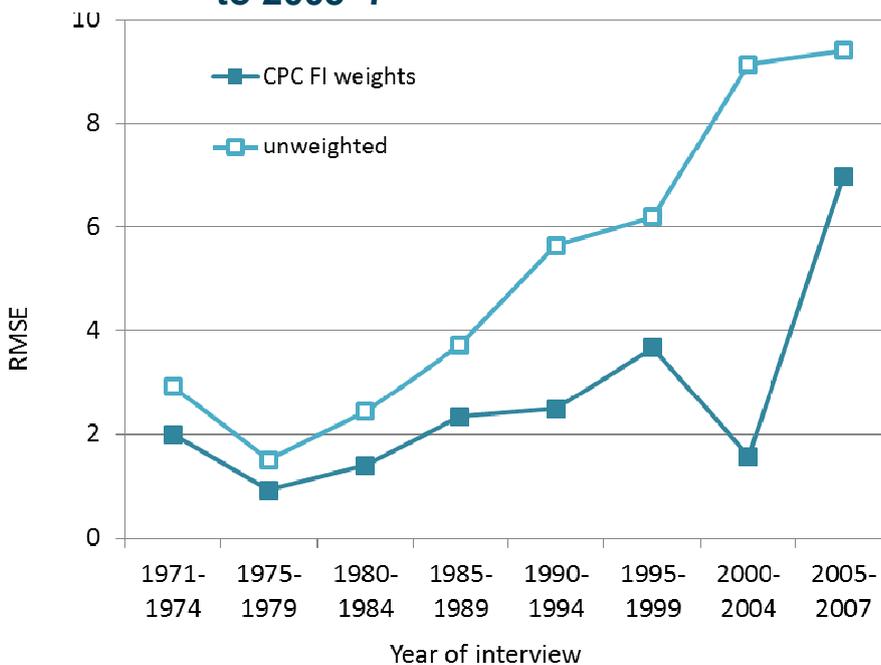
Note: Sample: women with a valid revised childbearing history

Figure 13 Ratio of age specific fertility rates based on revised fertility histories to national vital registration, survey years 1979–2007, weighted using CPC-FI weights. GHS, England and Wales, 1971–4 to 2005–7



Note: Sample: women with a valid revised childbearing history

Figure 14 Root mean square error*, unweighted and CPC-FI weighted age-specific fertility rates compared with vital registration, survey years 1979–2007. GHS, England and Wales, 1971–4 to 2005–7



Notes: *Root mean square error: the root of the mean squared deviation of the annual GHS estimates from the vital registration figure, i.e. root of the average squared bias.
 Sample: women with a valid revised childbearing history

Conclusion

In developing two new sets of weights for the GHS from 1979–2007, we have used the national level TFR as a criterion against which to evaluate the performance of our weights. Clearly, analysis of fertility and other aspects of the GHS demographic histories does not have, as a central goal, the estimation of the TFR, for which reliable national figures are available from vital registration statistics. But the satisfactory performance of our two sets of weights in bringing the GHS estimates of the TFR closely into line with vital registration figures suggests that the weights will perform well in analyses of the demographic histories collected in the FI section in general.

The utility of the CPC weights is not confined to demographic applications, as the household weights are applicable to analyses of GHS data of all kinds, allowing analyses of pooled series of the GHS to be weighted on a unified basis throughout, rather than having a mix of unweighted and weighted data sets before and after 1996. This unified weighting and the resulting potential reduction in bias creates more variable weights than the original ONS weights. More variable weights will feed in to standard errors of estimates, although a full evaluation of the impact is not straightforward both because we are pooling across years to estimate the period TFRs and because we do not have access to the cluster structure of the surveys. Further analysis could explore whether the use of constraints in the calibration, as is used by ONS, would reduce the variability in the weights beyond that achieved here by trimming the largest weights.

Acknowledgments

This research is funded by ESRC Grant number RES-625-28-0001. The Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with ONS and National Records of Scotland (NRS). The findings, interpretations, and conclusions expressed in this article are entirely those of the authors and should not be attributed in any manner to ONS or NRS. The General Household Survey is conducted by ONS. Access to the data is provided by the UK Data Archive²¹. The CPC GHS time series datafile on which this article is based was constructed in collaboration with Dr Ann Berrington and with the assistance of Mark Lyons Amos. We thank the Demographic Analysis Branch and the General Lifestyle Survey Branch of ONS for their help in clarifying various data issues and for providing the 2001 Census-GHS link adjustment factors. Our thanks also to Peter Smith for helpful advice.

Box 1 Glossary

CHAID Chi-squared Automatic Interaction Detector

This is a tree-like statistical procedure designed to identify groups of persons, defined by combinations of characteristics, so that the prediction of an outcome variable, non-response in the present case, is optimised.

Calmar

This is an acronym for 'Calibration to the Margin' and is a statistical program devised by INSEE, the French National Statistical Office. Calibration creates weights that reproduce a marginal distribution in the sample weighting classes that corresponds to the population distribution in those categories. Calmar often uses a method known as raking to achieve calibration.

GES Generalized Estimation System

This is a program developed by Statistics Canada that can implement complex survey estimation including model-assisted approaches such as the Generalised Regression Estimator (GREG). The GREG leads to calibrated estimators and is an alternative to raking but using an additive rather than multiplicative underlying model.

Normalization

Normalized weights are adjusted to the average weight in a given sample, and are such that the sum of the weights equals the sample size.

Trimming

Weights are trimmed by constraining them to an upper limit, the objective being to reduce the impact of applying the weights on the variance of the estimates.

Notes and References

- 1 Ali R, Binmore R, Dunstan S, Greer J, Matthews D, Murray L, Robinson S. 2009. *General Household Survey 2007*. London: Office for National Statistics, Appendix B.
- 2 Note, however, that while the vital registration figures are the best source of validation available for the fertility histories collected in the GHS, the two populations do not exactly coincide. Vital registration figures are based on births occurring during those years in GB, some of which may be to women who have died or emigrated. Conversely, the GHS figure, constructed retrospectively, includes births that may have occurred elsewhere during those years, including births to immigrants or to GB residents who had their births elsewhere and did not register them in GB.
- 3 Murphy (note 4) shows that fertility histories in recent rounds of the GHS have been quite inaccurate, with substantial over-statement of the level of childlessness, particularly among older women. However, work carried out at the ESRC Centre for Population Change has

established that the bias due to misreporting can be corrected to a considerable extent by supplementing the reported birth histories with information on own children in the household, recorded in the household schedule. Throughout this article, we use the best available fertility histories - up to 1993, those originally reported, and from 1994–2007 revised birth histories that include, together with the originally reported births, all the own children recovered from the household information section who were not reported in the FI section, and whose inclusion was consistent with the reported fertility history. We refer to these collectively as the 'revised fertility histories'. For further details see Ní Bhrolcháin et al (note 5).

- 4 Murphy M. 2009. Where have all the children gone? Women's reports of more childlessness at older ages than when they were younger in a large-scale continuous household survey in Britain. *Population Studies* 63: 115-133.
- 5 Ní Bhrolcháin M, Beaujouan É, Murphy M. 2011. Reported childlessness: A further look at cohort estimates based on survey time-series data. Centre for Population Change Working Paper 11, University of Southampton; and forthcoming in *Population Studies*.
- 6 Barton J. 2001. Developing a weighting and grossing system for the General Household Surveys in the United Kingdom. *Survey Methodology Bulletin* 49: 15-26.
- 7 A similar upward bias to survey fertility estimates due to higher non-response rates among childless women has been noted in European fertility surveys. See Festy and Prioux (note 8) and Kreyenfeld et al (note 9).
- 8 Festy P, Prioux F. 2002. *An Evaluation of the Fertility and Family Surveys Project*. New York and Geneva: United Nations.
- 9 Kreyenfeld M, Zeman K, Burkimsher M, Jaschinski I. 2011. Fertility data for German-speaking countries. What is the potential? Where are the pitfalls? MPIDR Working Paper WP 2011-003. Rostock: Max Planck Institute for Demographic Research. Available at: <http://www.demogr.mpg.de/papers/working/wp-2011-003.pdf>
- 10 The GHS was suspended in 1997 for a review and in 1999 for redevelopment prior to a relaunch in 2000.
- 11 Kim JK, Park M. 2010. Calibration estimation in survey sampling. *International Statistical Review* 78: 21-39.
- 12 Park M, Fuller WA. 2005. Towards nonnegative regression weights for survey samples. *Survey Methodology* 31: 85-93.
- 13 Särndal C-E. 2007. The calibration approach in survey theory and practice. *Survey Methodology* 33: 99-119.
- 14 Calibration consists in raking the adjustment factors using an iterative calculation, in order to adjust the weighted data to the structure and numbers of the GB population by age and sex, and by region; see Glossary and Battaglia et al, note 15.

- 15 Battaglia MP, Izrael D, Hoaglin DC, Frankel MR. 2004. Practical considerations in raking survey data. Paper presented at Annual Conference of the American Association for Public Opinion Research, Phoenix, Arizona. Available at:
www.abtassociates.com/presentations/raking_survey_data_2_JOS.pdf
- 16 Lemaître G, Dufour J. 1987. An integrated method for weighting persons and families. *Survey Methodology Bulletin* 13: 199-207.
- 17 Raghunathan TE. 2004. What do we do with missing data? Some options for analysis of incomplete data. *Annual Review of Public Health* 25: 99-117.
- 18 Little RJA. 1986. Survey nonresponse adjustments for estimates of means. *International Statistical Review* 54: 139-157.
- 19 Kreuter F, Olson K, Wagner J, Yan T, Ezzati-Rice TM, Casas-Cordero C, Lemay M, Peytchev A, Groves RM, Raghunathan TE. 2010. Using proxy measures and other correlates of survey outcomes to adjust for non-response: Examples from multiple surveys. *Journal of the Royal Statistical Society, Series A*: 173: 389-407.
- 20 Raking has less effect in adjusting the earlier GHS rounds than in later ones, presumably because the responding sample was more representative of the target population than in more recent years.
- 21 Office for National Statistics. Social and Vital Statistics Division. 2009. General Household Survey 1979-2007. In: Colchester Essex: UK Data Archive [distributor].

Appendix

Table A1 **Weighting classes for household non-response from CHAID analysis of 1991 Census-GHS link**

Level 1 split	Level 2 split	Level 3 split	Level 4 split	Weight class
Region North East Merseyside Yorks & Humbs W Midlands South East Scotland	No. of Cars 0 or 1	No. of dependent children 0 or 1	Household type 1 1 adult 16-59 Youngest 5-15 3+ adults no child	1
			Household type 1 2 adults 16-59 Youngest 0-4 2 adults, 1 or 2 60+ 1 adult only 60+	2
		No. of dependent children 2 or more		3
	No. of Cars 2 or more			4
Region North West E Midlands Eastern South West	Pensioner HH Pensioner in HH	SEG (grouped) Skilled Manual Partly-skilled manual Unskilled manual & others Not employed in last 10 years		5
			SEG (grouped) Professional Manager/employer Intermediate/jnr	6
	Pensioner HH No pensioner in HH	No. of adults 1		7
		No. of adults 2 or more	Social Class I II IV Not employed in last 10 years	8
			Social Class III _{nm} III _m V Other	9
Region London	Type of building Detached Semi-detached Terraced Converted flat/other			10
	Type of building Purpose built flat			11
Region Wales				12

Source: Barton, 2001

Table A2 **Weighting classes for household non-response from CHAID analysis of 2001 Census-GHS**

Figure D.A 2001 Weighting classes formed in the AnswerTree analysis

LEVEL 1 SPLIT	LEVEL 2 SPLIT	LEVEL 3 SPLIT	LEVEL 4 SPLIT	WEIGHT CLASS	
Region	Number of Rooms	Number of Pensioners in the household	Sex of the Household Reference Person		
West Midlands London	More than three	All pensioners	Female	1	
			Male	2	
		Two or two or more persons in the household but only 1 pensioner		3	
		Two or two or more persons in the household where more than one person is a pensioner No pensioners in the household		4	
	Three or fewer			5	
Yorkshire & Humberside East Midlands South West Wales	Household size				
	More than two			6	
	Two			7	
	One	Adults not employed			
One		One	8		
		Zero	9		
North East North West & Merseyside East of England South East Scotland	Number of Adults				
	More than two			10	
	Two	Accommodation Type			
		Purpose built flat Part of a converted or shared house Other			11
	Detached Semi-detached Terraced			12	
	One			13	

Source: Ali et al (2009), Appendix D

Table A3 **Weighting classes used for producing the GHS calibration weights, 1996–2003**

Age/sex	Region
0-4	London
5-14	Scotland
15-24 Male	Wales
15-24 Female	Other metropolitan
25-44 Male	Other non-metropolitan
25-44 Female	
45-64 Male	
45-64 Female	
65-74 Male	
65-74 Female	
75+ Male	
75+ Female	

Source: Barton, 2001

Table A4 **Weighting classes used for producing the GHS calibration weights, 2004–07**

Age/sex	Region
0-4	London
5-15	Scotland
16-24 male	Wales
16-24 female	Other Metropolitan
25-44 male	Other Non-metropolitan
25-44 female	South East
45-64 male	
45-64 female	
65-74 male	
65-74 female	
75+ male	
75+ female	

Source: Ali et al (2009), Appendix D.