Longitudinal and Life Course Studies 2011 Volume 2 Issue 3 Pp 260 - 285 *ISSN 1757-9597*

Cross-cohort changes in gender pay differences in Britain: accounting for selection into employment using wage imputation, 1972-2004

Jenny Neuburger
London School of Hygiene and Tropical Medicine
Jenny.Neuburger@lshtm.ac.uk
Diana Kuh
Medical Research Council National Survey of Health and Development, MRC Unit for Lifelong Health and Ageing
Heather Joshi
Centre for Longitudinal Studies, Institute of Education, University of London

(Received January 2011 Revised April 2011)

Abstract

This paper examines trends in the labour market position of British women and men from 1972 to 2004, using micro data from three British Birth Cohort Studies, of 1946, 1958 and 1970. Women's rates of employment and hourly pay have been lower than men's over this period, but generally increasing. Because employment decisions are influenced by the level of pay on offer, changes in women's relative pay may not be representative of changes in their labour market position. We accounted for selection into employment by imputing missing hourly wages for non-employees using observed wages of employees of the same sex and age with similar work and family histories, matched on their propensity score. At each survey, women's median hourly pay was lower than men's. Although relative pay increased across the cohorts, it decreased with age within each cohort. Accounting for selection into employment gave a lower estimate of young women's potential pay relative to men's in the two earlier cohorts. This evidence supports the view that the improvement in young women's labour market position since the 1970s has been substantial, and is underestimated in pay trends for the working population.

Keywords: Wages, gender gaps, employment, British Birth Cohorts, sample selection, imputation, propensity score matching

1. Introduction

A broad measure of women's and men's relative power and status in the workforce is the ratio of their average hourly pay. Not only does this measure the differences in pay for the same types of work, but also reflects the level of skill and responsibility involved in the jobs that they do. Thus, it is a measure of inequality in the structure of the workforce. However, it is a potentially biased measure because it excludes the non-working population, who may be out of work precisely because the wages on offer are too low. In particular, pay comparisons within the workforce tend to exclude a higher fraction of women than of men with low potential pay. The importance of such selection biases in wage comparisons was raised in the labour economics literature by Gronau (1974) and Heckman (1977) and has been the subject of more recent studies (Blau and Kahn 2006; Blundell et al 2007; Olivetti and Petrongolo 2008).

We focus on gender pay ratios, without making any adjustment for factors such as education and employment experience, although such adjustment is often done in the literature on gender pay discrimination (for example by Wright and Ermisch

1991; Joshi et al 2007). From a life-cycle perspective, this unadjusted ratio can be seen as a measure of the cumulative effects of pay inequality, since different rates of pay for women and men may affect investment in education and the gender division of domestic work, with cumulative impacts on productivity as well as pay. From a historical perspective, the broad ratio reflects changing institutional arrangements and employment opportunities. It is particularly pertinent to look at the overall gender ratio across the period studied, when the introduction of equal opportunities legislation in the 1970s had a major impact. To reveal its full effect, it is necessary to account for selection biases in unadjusted gender pay ratios. The estimation of adjusted gaps addresses a different question (i.e. differential pay conditional on education and experience) and is the subject of a separate investigation (Neuburger 2010) which relies on the findings in this paper.

Accounting for selection biases in wages is important for the period of British history we focus on, which saw major changes in women's and men's employment and pay. The implementation of the Equal Pay Act in 1975 led to increases in women's relative pay, following decades of no change (Zabalza and Tzannatos 1985; Manning 1996). The introduction and extensions of maternity leave after 1975 also contributed to an increase in mothers' employment across the wage distribution, apart from the lowest paid (Gregg et al 2007). Increases in women's relative wages in turn contributed to a decline in births and an increase in female employment in the 1970s, whilst rising pay was both a cause and effect of rising relative female education (Joshi et al 1985; Ermisch 1988). For men, decreases in economic activity in the 1980s and 1990s were concentrated amongst unskilled groups facing a drop in relative wages (Disney and Webb 1991; Faggio and Nickell 2003). Whereas women's employment rate increased from 56% to 70% of the working-age population between 1971 and 2008, men's employment rate decreased from 91% to 80% (Labour Force Survey 1971-2008).

Age comparisons of gender pay differentials may also suffer from selection biases, owing to more continuous life-cycle employment of higher-waged women around childbearing years. Wage-age profiles estimated for quasi-cohorts in the New Earnings Survey (supplementary analysis in Manning and Swaffield 2008) and General Household Survey (Harkness 2005) show a decrease in women's average pay relative to men's up to the age of 40, alongside cross-cohort increases. However, some of the decrease in women's earning power immediately after having children may not be measured in pay trends, owing to positive selection back into employment.

Several studies have found evidence of substantial selection biases in gender wage comparisons across different time periods and countries, arising from differences in employment (Blau and Kahn 2006; Blundell et al 2007; Olivetti and Petrongolo 2008). Focusing on a similar question to ours, Blundell et al (2007) looked at changes in wage differentials in the UK between 1978 and 1998 using the Family Expenditure Survey (FES), with and without controlling for selection biases. They found that selection into employment masked some of the improvement in the position of women in work.

In this paper, we present newly compiled evidence from the British Birth Cohort Studies covering three decades at the end of the 20th Century. Our working hypothesis was that low wage opportunities for women in the earliest cohort in the 1970s were partly masked by low rates of employment, combined with positive selection into employment. Thus, changes in women's labour market position may not be fully represented in pay trends for the working population. We extend the analysis of Blundell et al (2007) in three main ways. First, we cover a longer period, starting in 1972, before the implementation of equal opportunities legislation, and ending after the Millennium (2004). Second, we are able to draw some distinctions between cohort and life-cycle effects by using large samples from three birth cohorts at three different ages. Third, we use detailed data on individual childhood, work and family histories in an alternative method of controlling for selection into employment.

Section 2 describes the dataset. Section 3 describes the wage imputation method used. Sections 4 and 5 present our results and conclusions.

2. Data

We compare three of the British Birth Cohort Studies. These are ongoing national multi-purpose studies of individuals born in March 1946, March 1958 and April 1970: the Medical Research Council

National Survey of Health and Development (NSHD); the National Child Development Study (NCDS); and the 1970 British Cohort Study (BCS70). Each of the studies took as its original sample all of the British births over a selected week. None of the three studies are fully representative of the adult population living and working in Britain, since each excludes some or all immigrants. For a profile of each study, see Wadsworth (2006) and Kuh et al (2011), Power and Elliott (2006) and Elliott and Shepherd (2006). For a comparison of the three cohorts, see the volume edited by Ferri et al (2003).

The three studies, taken together, offer important attractions for the analysis of employment and pay trends in Britain since the 1970s. A key advantage is that the 1946 cohort had a distinctly different history to the two later cohorts. Not only entering and anticipating the labour market before the Equal Pay Act, women in the 1946 cohort grew up in an era when girls did not equal male achievements in education and were typically not expected to combine parenthood with careers. Wage estimates for the 1946 cohort provide critical baseline estimates for cohort change.

A second major advantage is the detailed longitudinal information on childhood, education, employment and family histories. This makes possible the joint analysis of employment, family and wage data. A disadvantage is that earnings data were not collected frequently, not at the same ages nor in the same form for each cohort, in part owing to the somewhat different aims of the studies and in part owing to funding constraints and opportunities at different points in time. The data are not well suited to the longitudinal analysis of wage dynamics, unlike the British Household Panel Survey (BHPS), for example, but do allow the study of changes experienced by the whole cohort at different points over the life-cycle (Table 1).

Table 1. Description of surveys collecting wage data from the birth cohorts

Birth cohort	Survey year	Age	Survey method	Sample (men)	Sample (women)	Response rate *
1946	1972	26	home interview	1,897	1,853	85%
1946	1977	31	postal questionnaire	1,668	1,672	74%
1946	1989	43	home interview	1,635	1,627	80%
1958	1981/82	23	home interview	6,268	6,271	76%
1958	1991	33	home interview	5,630	5,836	72%
1958	1999/2000	42	home interview	5,627	5,794	73%
1970	1996	26	postal questionnaire	4,101	4,902	55%
1970	1999/2000	29/30	home interview	5,461	5,784	70%
1970	2004	34/35	home interview	4,625	5,039	61%

**The longitudinal response rate defined as the percentage of the target sample who participated at each survey, excluding individuals known to have died or emigrated. Permanent refusals were included in the denominator, giving slightly lower estimates than those presented in Wadsworth et al (2003) for the 1946 cohort. Figures for the 1958 and 1970 cohorts are estimates from Plewis et al (2004) and Ketende et al (2010).*

Derived variables

A major exercise was undertaken to derive comparable variables for the cohorts and to check and edit the data (see Neuburger 2010). Definitions are given in Appendix Table A1.

Employment status: Cohort members were asked to describe their current main activity at each survey. Measures of part-time work are based on self-defined part-time status, although questions were accompanied in most of the surveys by a prompt defining part-time work as working less than 30 hours a week.

Gross hourly earnings: Before-tax hourly earnings are used as the measure of pay in our analysis. Income from self-employment was excluded from our analysis, since it is organised, declared and measured differently from income from employment, and represents a return on assets and enterprise as well as labour. For employees, we calculated hourly earnings in two steps: first, weekly earnings were calculated from responses on before-tax pay and corresponding pay periods; second, hourly earnings were calculated by dividing the weekly figure by reported weekly hours of work. Overtime work was included in the numerator (pay) and denominator (hours). Measures were adjusted to January 2000 prices using the Office for National Statistics long-term indicator of prices.

Employment experience: Estimates were made of the number of years spent in employment and the number of years in full-time and part-time work. For the 1958 and 1970 cohorts, these figures were derived from retrospective data on job and unemployment histories collected from cohort members at each adult survey.

For the 1946 cohort, the job history data were not sufficiently complete for women to create full work histories. Instead, for women and men, estimates were made of number of years in work from the age of 25, rather than since leaving full-time education. This may affect the quality of wage imputations at age 26. Including an existing variable containing the total number of months spent out of work between the ages of 18 and 25 in our imputation models did not alter our results.

Social class of first job: For the 1958 and 1970 cohorts, information was included on the social class status of the first job held after leaving full-time education.

Highest educational qualification: Measures of the highest qualification held at each age include academic and vocational qualifications obtained in adult life. These were grouped using a classification devised by Makepeace et al (2003). We refer to the categories by the main academic exams taken in the English school system over this period: O-levels were basic academic qualifications, usually taken at age 16; A-levels were advanced academic qualifications, usually taken at age 18; diplomas were below-degree qualifications, including some teaching, nursing and lower-level professional qualifications; degrees cover undergraduate qualifications and higher.

Number and ages of children: Variables on the number and ages of children in the household were derived from retrospective birth histories for women and from information on household composition. Birth history data were used for ages up to 43 in the 1946 cohort, age 42 in the 1958 cohort and age 34 in the 1970 cohort.

Childhood mathematics scores: Scores from mathematics tests taken at age ten for the 1970 cohort and age 11 for the 1946 and 1958 cohorts were used as indicators of educational achievement at these ages. Standardised scores were calculated for the full sample (girls and boys) who took the tests.

Childhood variables: Variables containing information on cohort member's family size in childhood, their father's social class and mother's and father's ages and schooling were also included, also having been previously established as predictors of future earnings and employment in the birth cohorts (e.g. Kuh and Wadsworth 1991; Kuh et al 1997).

Individuals with missing items for highest qualification, employment experience or, for women only, numbers and ages of children, were excluded from our analysis. Cases with missing maths or childhood variables were included in the models. Missing maths scores were imputed from other childhood variables and dummy variables were included to indicate a missing item for maths scores and other childhood variables. Missing wages for employees were imputed using the full set of covariates, using the same methods as described below for imputing non-employee wages. The results were also robust to their exclusion from the samples and they were excluded from sample sizes on which standard errors were based (see below).

Effects of sample stratification and survey non-response

For the 1946 cohort, the sample followed up at age two included 5,632 children of the 13,687 births in the original maternity survey (91% of births in one week in March 1946 in England, Scotland and Wales). Babies born to unmarried mothers (N=672) and multiple births (N=180) were excluded from the study. All children born to fathers in non-manual and agricultural occupations were included in the age two survey, but only one in four born to fathers in urban, manual occupations, the aim being to preserve roughly equal numbers from the different social classes given funding constraints. We found that class origins were strongly associated with future earnings and that these effects differed by gender. A weighting variable is provided with the datasets, taking the value 4 for individual cases representing four cohort members and taking the value 1 otherwise. The variable was used as a covariate in the probit models used to estimate propensity scores (see below) and as a weight to estimate summary statistics.

Adult response rates decreased across the cohorts and also after their twenties within the two earlier cohorts (Table 1). Response rates to the 1996 postal survey (age 26) of the 1970 cohort were particularly low (55%), partly because of limited time and resource to trace cohort members (Plewis et al 2004). Studies of longitudinal non-response in the studies show that individuals who left the studies were more likely to be male and to have experienced disadvantage in childhood (Wadsworth et al 2003; Hawkes and Plewis 2006).

We undertook a supplementary analysis to characterise and quantify biases in our wage data arising from attrition and other survey non-response. We used the mathematics scores for cohort members from ages 10 or 11. Individual scores were strongly positively correlated with future earnings and with the probability of survey response in adulthood. As such, they are useful indicators of wage biases arising from non-response. We found that mean scores were generally higher amongst respondents than amongst non-respondents by roughly a third of a standard deviation (Appendix Table A2). These differences are comparable in size to those associated with selection into employment (Appendix Table A6).

We also compared wage ratios and employment rates in the cohort samples to those estimated from the Family Expenditure Survey (FES) for the same years (Table 2). Our FES cross-section samples were restricted to individuals either the same age or one year older or younger than cohort members. The broad pattern of cross-cohort and within-cohort trends in gender pay ratios was mirrored in the FES estimates and the confidence intervals on the estimates were overlapping in all cases. There was some evidence of upward bias in cohort ratios at age 43 in the 1946 cohort, at age 42 in the 1958 cohort and at age 26 in the 1970 cohort. However, the size of this bias is unclear since the FES estimates were based on comparatively small sample sizes and are imprecise i.e. with wide confidence intervals. There was also evidence that estimates of employment rates were slightly upward biased in the 1989 (age 43) survey of the 1946 cohort and in all three surveys of the 1970 cohort, compared to the FES samples. We consider the potential impact of these patterns of survey non-response on our conclusions.

Table 2. Median hourly wages and female-to-male ratios: comparison with FES

Survey year (age)	Sample size cohort	FES	Median wage, £ (95% CI) cohort	FES	Female-to-male ratio (95% CI) Cohort	FES
1946 cohort						
1972 (26)	2,396	592	5.4 (5.3, 5.5)	5.3 (5.1, 5.5)	0.68 (0.64, 0.72)	0.68 (0.63, 0.74)
1977 (31)	2,082	663	5.6 (5.5, 5.8)	5.6 (5.5, 5.8)	0.64 (0.61, 0.67)	0.65 (0.61, 0.69)
1989 (43)	2,291	605	7.0 (6.7, 7.2)	6.9 (6.4, 7.4)	0.60 (0.55, 0.64)	0.53 (0.48, 0.58)
1958 cohort						
1981 (23)	8,629	572	5.2 (5.1, 5.2)	5.2 (5.1, 5.4)	0.84 (0.82, 0.85)	0.79 (0.73, 0.85)
1991 (33)	7,726	535	7.6 (7.5, 7.7)	7.1 (6.8, 7.4)	0.69 (0.67, 0.72)	0.63 (0.56, 0.70)
2000 (42)	8,213	494	7.9 (7.8, 8.0)	8.2 (7.7, 8.8)	0.67 (0.65, 0.68)	0.58 (0.51, 0.65)
1970 cohort						
1996 (26)	6,521	450	6.7 (6.6, 6.8)	6.7 (6.4, 7.1)	0.89 (0.87, 0.91)	0.81 (0.73, 0.89)
2000 (30)	8,258	542	7.6 (7.5, 7.7)	7.8 (7.4, 8.3)	0.87 (0.84, 0.89)	0.81 (0.71, 0.92)
2004 (34)	7,048	533	8.7 (8.6, 8.8)	8.7 (8.2, 9.2)	0.80 (0.77, 0.82)	0.87 (0.77, 0.97)

Notes: Cohort samples include employees with observed wages and imputed missing wages. Samples from the Family Expenditure Survey (FES) include individuals either the same age, or one year older or younger than the birth cohort samples. 95% confidence intervals are shown in brackets.

3. Methods

For individuals who were not in paid work, the potential wage was defined as the wage that they could expect to earn if they entered work. We use the term "potential wage", rather than the traditional economic concept of the "wage offer", since it is not linked to one theoretical model of the labour market (Olivetti and Petrongolo 2008) and more naturally encompasses situations in which individuals are not seeking work, have been out of work for a long period and have only a vague expectation of what they could earn if they did get a job. Wages were not imputed for self-employed individuals, based on the assumption that their labour earnings were missing at random. This assumption was supported by a comparison of their characteristics, relative to those of employees. Our conclusions were also robust to their inclusion.

Since potential wages of non-workers are unobserved, statistical methods to estimate these inevitably draw on additional modelling assumptions, such as:

- *Selection on observables.* The potential wages of non-workers are, on average, the same as observed wages of workers with the same observed characteristics.
- *Selection on unobservables, plus exclusion restriction.* The potential wages of non-workers are the same as wages of similar workers, but that determinants of the employment decision, uncorrelated with potential wages, can be used to estimate the unobserved selectivity bias (Gronau 1974; Heckman 1979).

- *Positive selection.* Potential wages of non-workers are lower, on average, than observed wages of workers with the same characteristics (Blundell et al 2007).
- *Constancy of individual relative wage over time.* An individual's relative wage remains fairly constant over time and their potential wage is, on average, the same as their observed wage on a previous or future occasion when in work (Blau and Kahn 2006; Olivetti and Petrongolo 2008).

Which assumption is used depends on the data available, as well as on the theoretical stance taken on credibility of different assumptions. The assumptions can also be formulated in weaker terms (Manski 1989; Blundell et al 2007).

For women, employment participation tends to be positively correlated with their own wage prospects, based on own levels of education, age at having a first child, employment experience and past wages, but negatively correlated with their partner's income (Joshi 1986; Gregg et al 2007). Historically, partner's income and other family structure variables have been used as types of instrumental variable to quantify the likely size of unobserved selection biases in wages (Gronau 1974; Heckman 1977).

Using family structure and health variables as exclusion restrictions in Heckman's two-step model, unobserved selectivity biases in wages have previously been estimated to be small and non-significant for the 1946 and 1958 cohorts (Joshi and Paci 1998; Kuh et al 1997). Such instruments are less credible for later cohorts, such as the 1970 cohort, since there evidence of increasingly strong correlations between characteristics of spouses in their social origins and levels of education (Blossfield and Timm 2003). More credible exclusion restrictions are those based on administrative or structural arrangements that affect employment, such as out-of-work benefits entitlement (Blundell et al 2007), but this type of information was not collected in the birth cohorts.

Wage imputation method

We have used a wage imputation method which relies on the assumption of selection on observables. The detailed information on individual work and family histories available in our data was an important pre-condition for this assumption. We also exploited the longitudinal aspect of the wage data to test the sensitivity of our results to alternative assumptions (see below).

Imputation methods come from statistical work on methods to handle bias arising from missing data in surveys (see Little and Rubin 2002). These methods have been further developed for handling missing data in large-scale Government surveys. We have used a form of nearest-neighbour imputation based on propensity score matching (Rosenbaum and Rubin 1983).

Our method involved two steps. First, a propensity score was estimated using a probit model with a binary response, taking the value 1 for non-employees and 0 for employees (Appendix Tables A4 and A5). The second step was to match each non-employee to a potential wage donor, i.e. individuals in work despite a low probability of working, based on their propensity score. For each matched pair, the missing value was replaced with the value of the observed wage. The estimation of the probit model and the nearest-neighbour matching were carried out together using the psmatch2 program in Stata (Leuven and Sianesi 2003). We included a common support restriction which excluded individuals with a propensity score outside the range of scores for the opposite group.

A summary of differences in mean maths scores and other model statistics before and after matching are given in Appendix Table A6. The difference before matching gives an indication of the strength of biases arising from positive selection into employment, based on observed characteristics. The difference after matching reflects the degree of similarity between matched samples, and the accuracy of imputed values for non-workers. Sample sizes are given in Appendix Table A7.

Summary statistics and standard errors

The ratio of women's to men's median hourly wage is the summary statistic used as the measure of relative pay. The ratio of women's to men's median potential hourly wage, including imputed values for non-workers, is the measure of relative pay opportunities. Methodologically, comparing medians places less reliance on the goodness of individual imputations. Substantively, cross-cohort comparisons of relative median wages, rather than relative mean wages, places less weight on the differential effects on women and men of rising wage inequality.

Standard errors on these ratios were estimated using bootstrap methods, rather than analytically, since there is no formula for ratios of medians. The size of samples used in the bootstrap replications was restricted to the size of the original non-imputed sample with observed wages.

Sensitivity analysis

We exploited availability of longitudinal wage data to test the robustness of our results to the assumption of "selection on observables". Separately for women and men who were not in work, we assigned their missing wage above or below the observed median wage, based on the position of their last or next observed wage. These assignments capture fixed, unobserved selectivity biases, but assume that individuals experience a limited amount of individual life-cycle wage mobility. The results of this exercise were used as the basis for a sensitivity analysis; applying the fractions of below-median wages, estimated from the sample of non-working individuals with observed wages at a previous or later survey, to the whole non-working population and re-estimating gender pay ratios.

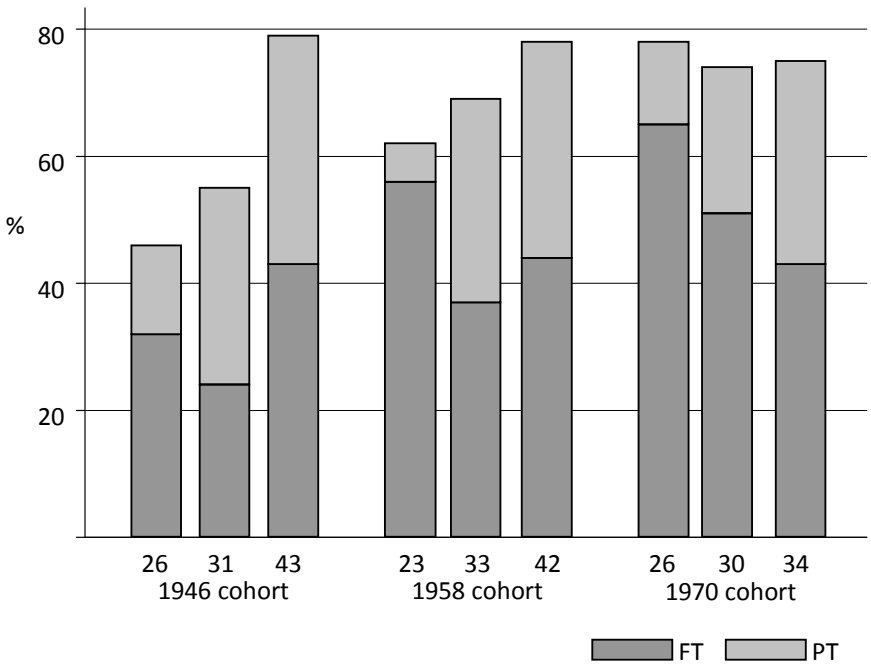
4. Results

Overview of trends

For women and men in the 1958 and 1970 cohorts, the expansion of basic and higher educational opportunities delayed the start of work and changed patterns of earnings and employment. More than 60% of the 1946 cohort left school at ages 15 (then the minimum school leaving age) or 16 and more than half of the cohort had either no or very low formal qualifications i.e. not having either a General Certificate of Education (GCE) `O'-level, a Certificate of Secondary Education (CSE) grade 1, a Scottish Standard grade 1 or a vocational equivalent, which are exams usually taken at 16. Around 60% of the 1958 cohort left at age 16 (by then the minimum leaving age), and around a quarter of the 1958 cohort and a fifth of the 1970 cohort had no or very low formal qualifications.

The change in women's employment across the cohorts at younger ages is equally striking (Figure 1, Appendix Table A3). Just under half of women in the 1946 cohort were in paid work at age 26 in 1972, compared to nearly 80% in the 1970 cohort at the same age quarter of a century later (1996). The fraction of women in full-time work also increased across the cohorts at younger ages, but by their mid-thirties, at least 40% of employed women were working part-time in all three cohorts.

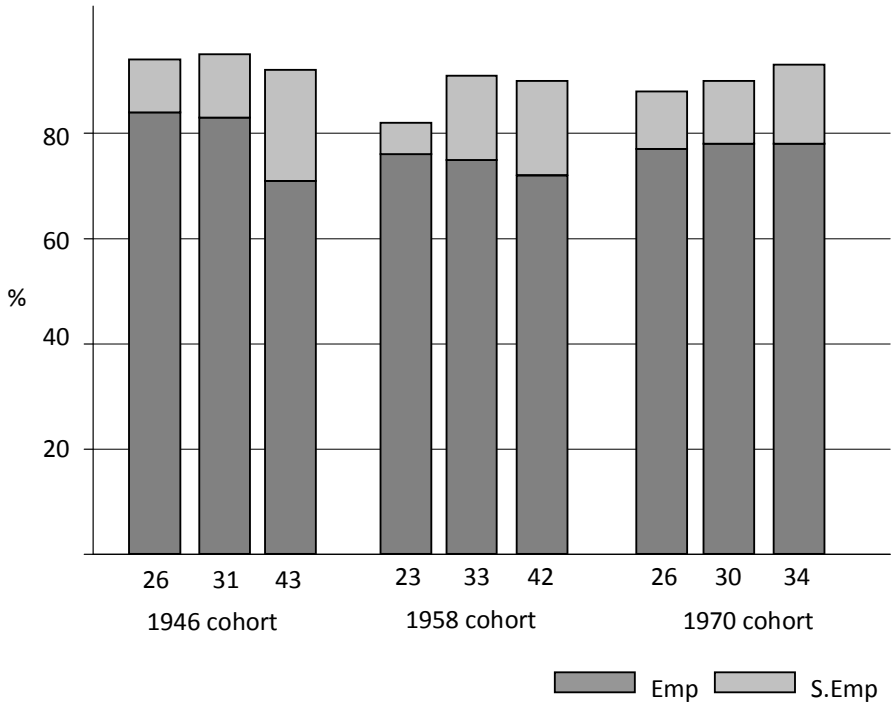
Figure 1. % of women in full-time or part-time work, by age and cohort



Full-time and part-time work here includes self-employed.

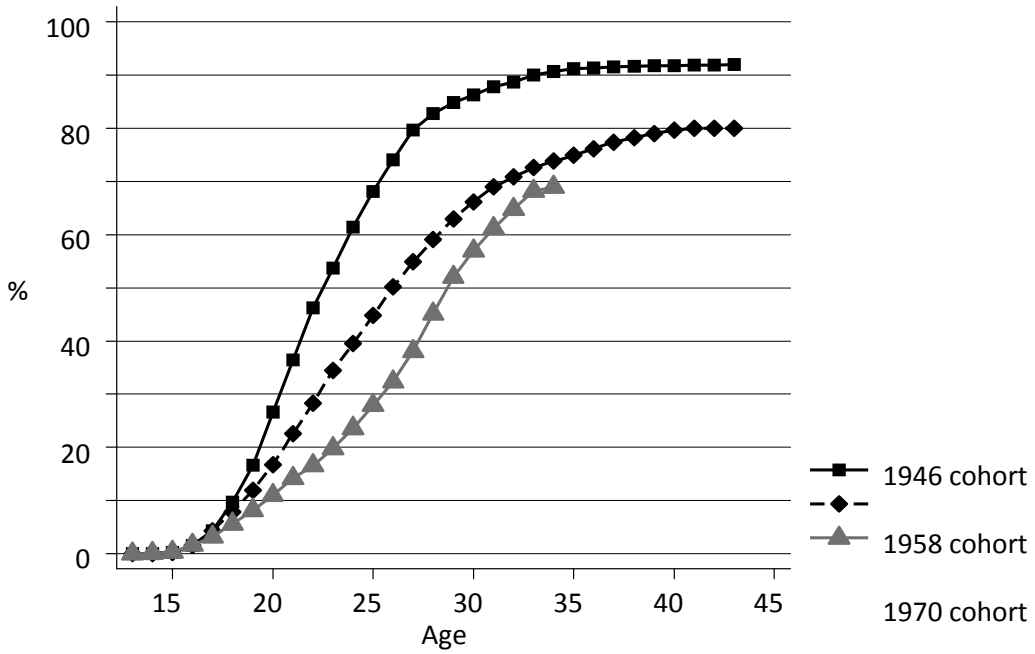
In contrast to women, men's rates of employment decreased slightly across the cohorts at young ages, although they remained high (Figure 2, Appendix Table A3). The decrease was largest between the 1946 and later cohorts when in their twenties and thirties, with the two later cohorts affected by recessions at the start of their careers. Moreover, the employment rate for the age 26 sample of the 1970 cohort is likely to be an overestimate, owing to the low response to the postal survey.

Figure 2. % of men in employment and self-employment, by age and cohort



The cross-cohort increase in women's rate of employment was composed both of an increase in the proportion of women without children and an increase in rates of maternal employment. Years of childbearing became more spread out and, on average, later for women in the 1958 and 1970 cohorts (Figure 3), with effects on the composition of female workforce at different ages. The proportion of mothers working a year after a first birth also increased from a fifth in the 1946 cohort to nearly two-fifths in the 1958 cohort and close to 60 per cent in the 1970 cohort.

Figure 3. % of women who were mothers by each age, by cohort (live births only)



The timing of childbirth differed by social and educational status in all three cohorts. More educated women were likely to have had their children later in life. For the 1946 cohort, there were strong social and educational differences in the timing of childbirth over a relatively small range of ages; from the early-twenties to the early-thirties. By age thirty-one, most women in the 1946 cohort had become mothers and there were few systematic differences between the groups in and out of employment at this age. For the two later cohorts, the compositional effects of delays in childbirth amongst more highly qualified women were mostly offset by their more rapid return to work.

Across the cohorts, women's median pay increased relative to men's at all ages (Figure 4). In their twenties, the ratios for the 1946, 1958 and 1970 cohorts were 0.68, 0.84 and 0.90 respectively. In their thirties, the respective ratios were 0.62, 0.70 and 0.86/0.80. In their early forties, the change across the 1946 and 1958 cohorts was less, with a ratio of 0.60 (95% CI 0.57, 0.63) for the 1946 cohort in 1989 and 0.67 (0.65, 0.69) for the 1958 cohort in 1999/2000 (Table 3). These patterns are also consistent with evidence of decreases in women's relative pay with age (Manning and Swaffield 2008), here including the impact of moves into lower paid part-time work (Connolly and Gregory 2009).

Changes in the relationship between wages and non-employment

Nearest-neighbour imputation provides a pseudo distribution of wages for non-employees, which can be used to understand where in the wage distribution non-employment is occurring for women and men at different ages and how this has changed across the cohorts.

For women, median imputed wages for non-employed women were systematically lower than observed wages of employed women (Appendix Table A8). Those out of work were also consistently over-represented in the bottom quartile of the observed wage distribution (Appendix Table A9). At age 26, over 70% of non-employed women in the 1946 cohort had imputed wages lower than the observed median wage. However, this pattern changed with age, and by age 43, the imputed wage distribution was bimodal; non-employed women were slightly over-represented in the top quartile of the observed wage distribution, as well as in the bottom quartile. For the 1958 and 1970 cohorts, women's wages were consistently over-represented in the lower part of the observed wage distribution and under-represented in the upper part (Appendix Table A9).

For men, median imputed wages for non-employed men were also systematically lower than median observed wages of employed men (Appendix Table A8). For men in the 1946 cohort, imputed wages for the non-employed were also over-represented in the lower part of the wage distribution, although they were also over-represented in the top quartile at age 26 (Appendix Table A9). In the two later cohorts, imputed wages of non-employed were also heavily concentrated in the lower part of the observed distribution, and appeared to become more so with age. An exception was at age 26 in the 1970 cohort, when non-employment seemed to occur more evenly throughout the wage distribution.

Population estimates of women's and men's median potential pay

For women, the inclusion of imputed wages of non-employees in population samples had the greatest quantitative impact on the population median for the 1946 cohort at age 26, since half of women were out of work. In all three cohorts, selection biases in women's wages occurred mainly around childbearing ages in each cohort; most strongly in evidence at age 26 in the 1946 cohort; at age 33 in the 1958 cohort; and at age 30 and 34 in the 1970 cohort. However, the impact of selection biases decreased across the three cohorts as the fraction of women out of work decreased from a half, to a third, to just under a quarter (Appendix Table A3).

For men, the inclusion of low imputed wages for non-employees in population samples did not have an impact on estimated median wages, since they comprised a small fraction of the population across the cohorts at all ages (Appendix Table A3).

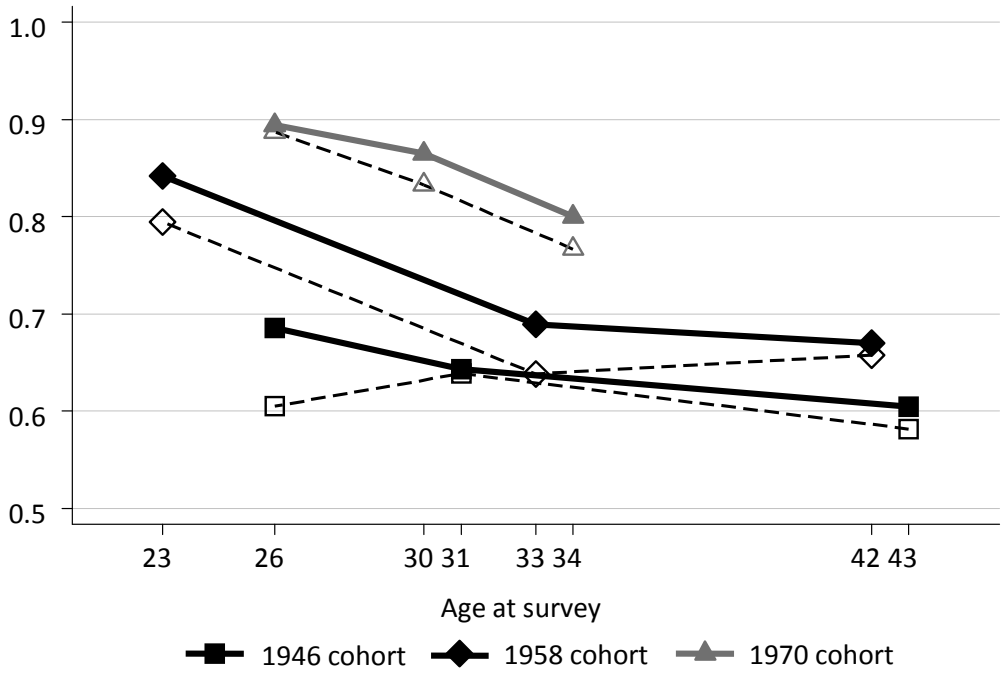
Putting these trends for women and men together, the aggregate picture is of a cross-cohort increase in women's median potential pay, relative to men's, that is understated in median pay ratios for employees (Figure 4). The 95% confidence intervals on estimated employee and population sample ratios are non-overlapping for the 1946 cohort at age 26, for the 1958 cohort at age 23 and for the 1970 cohort at age 30 (Table 3).

Table 3. Female-to-male ratios of medians, by survey

Birth cohort	Survey year	Age	Employee sample ratio (95% CI)	Population sample ratio (95% CI)
1946	1972	26	0.68 (0.64, 0.72)	0.61 (0.58, 0.61)
1946	1977	31	0.64 (0.61, 0.67)	0.64 (0.60, 0.67)
1946	1989	43	0.60 (0.55, 0.64)	0.58 (0.55, 0.61)
1958	1981/82	23	0.84 (0.82, 0.85)	0.79 (0.78, 0.81)
1958	1991	33	0.69 (0.67, 0.71)	0.64 (0.62, 0.66)
1958	1999/2000	42	0.67 (0.65, 0.68)	0.66 (0.64, 0.68)
1970	1996	26	0.89 (0.87, 0.91)	0.89 (0.87, 0.91)
1970	1999/2000	29/30	0.86 (0.84, 0.89)	0.83 (0.81, 0.85)
1970	2003	34/35	0.80 (0.77, 0.82)	0.77 (0.74, 0.79)

Notes: Employee samples include observed and imputed missing wages. Population samples include observed wages, plus imputed missing wages and imputed potential wages for non-employees.

Figure 4. Female-to-male ratios of median observed and potential pay, by survey



Notes: The solid thicker lines show the female:male ratio of median pay for employees. The dashed lines show the female:male ratio of median potential pay for the whole samples, including employees and non-employees.

Sensitivity analysis

We exploited the longitudinal aspect of the wage data to test the sensitivity of our estimates to alternative imputation methods. Our nearest-neighbour imputation method relied on the assumption of "selection on observables". An alternative assumption is that the position of an individual's wage or potential wage relative to the median remains the same over time, compared to those of the same sex and cohort. This assumption is supported by evidence for employees with repeat wage observations; around 70%-80% of individuals, both women and men, had wages in the same position relative to the median at consecutive surveys. We did not use this as a main imputation strategy because only between 30% and 60% of the non-employed samples had an observed wage at a subsequent or previous survey (see sample sizes in Appendix Table A10.)

Instead, for these restricted samples, we compared the fraction of imputed wages for non-employees that fell below the observed median wage using nearest-neighbour imputation with the fraction based their observed wage at another survey. There was evidence of position selection into work for women around childbearing ages from both methods (Appendix Table A10). However, for the 1946 cohort women, the fractions of below-median wages are lower when imputed from a future observed wage, rather than the nearest-neighbour wage. This may be owing to unobserved selectivity bias in the nearest-neighbour imputations, but may also be owing to a genuine recovery in the wage position of non-working mothers relative to women who had children at a later age. Applying this lower fraction of below-median wages to the complete non-employed female sample did not alter our conclusions.

5. Discussion

Summary and study limitations

Using newly assembled data from three British Birth Cohort Studies, we showed substantial cross-cohort increases in women's median pay, relative to men's, alongside decreases with age within each cohort. These findings confirm estimated trends for quasi-cohorts derived from cross-sections of the General Household Survey and New Earnings Survey (Harkness 2005; Manning and Swaffield 2008).

Accounting for selection into employment, we found evidence that the unequal position of young women in the labour market in the early 1970s was underestimated in the relative pay of those who were employed. Around half of women in the 1972 sample were not in work, and their estimated potential wages tended to be lower than those of women in work. However, as female employment rates increased across the cohorts, wage biases associated with positive selection into employment decreased. Taken together, these results suggest that the cross-cohort improvement in women's labour market position is underestimated in changes in relative pay for the working population. This is consistent with estimates for 1978-1998 based on the Family Expenditure Survey (Blundell et al 2007).

Over the life-cycle, selection biases occur around childbearing for women: in their 20s for the 1946 cohort and in the early 30s for the two later cohorts. In general, the decline in the relative potential pay of women around childbearing years appears to be partly masked by positive selection into employment. However, in the 1946 cohort, our wage observations do not go back far enough to say whether decreases occurred after childbirth or whether similarly low pay was a feature of women's work before having children. Although the wage position of non-employed women, relative to employed women, appeared to deteriorate in the two later cohorts, selection effects had decreasing impacts on the population median because a decreasing fraction of women were out of work at any one point in time.

Our main method of accounting for selection into employment was based on the assumption that non-working women and men could, on average, expect to earn a similar wage to working women and men who had similar levels of qualifications, employment experience and numbers of children. This is a simplifying assumption and, in reality, the reasons for working or not working are complex, depending on own wage prospects, employment sector, family circumstances, childcare costs and individual preferences. Systematic differences in the wages of those in and out of work may arise from differences in individual or employer characteristics, which are either not measurable (such as individual motivation) or not measured in our dataset (such as employment sector). As a check on the robustness of our results, we used longitudinal data to assign non-working individuals a potential wage above or below the median based on any observed wage at a previous or future survey. This alternative method also indicated positive selection into employment at younger ages and our conclusions were not changed.

A limitation of our data is that survey samples are not all representative of the cohort, owing to refusals to participate, or, more often, cohort members not being traced at the time of the survey. We found some evidence of bias in wage estimates arising from non-random non-response. Comparing female-male pay ratios in the cohort data to those based on small comparable cross-sections from the Family Expenditure Survey (FES), the general pattern of cross-cohort increases, and within-cohort decreases, was the same. There was some evidence that the ratio was upward biased at the age 43 survey of the 1946 cohort, at the age 42 survey of the 1958 cohort and at the age 26 survey of the 1970 cohort. As a consequence, we may slightly underestimate within-cohort decreases in women's relative pay in the two earlier cohorts into their forties, and overestimate cross-cohort increase for the 1970 cohort at age 26. We may also slightly underestimate the impact of selection bias in the 1970 cohort at this age, since the survey may have had lower response rates among the non-employed.

Conclusions

Debate continues about the causes of, and justification for, women's lower rates of hourly pay, compared to men's. There is a large literature focusing on individual causes, and separating the effects of gender on pay from the effects of having children, working part-time, spending periods out of work, and variations in education, to which this material is now placed to contribute. However, a broad measure of inequality in the structure of the labour market is important in its own right for historical comparisons, since individual decisions are shaped and constrained by institutional arrangements.

For women born in 1946, three decades before the implementation of the Equal Pay Act, relatively low rates of pay and long periods out of work to raise children remained the social norm into the 1970s. There is strong evidence of positive selection into the workforce. Arguably, low rates of pay may have created disincentives to education or

employment, but it is hard to establish the direction of causation. Also, the overrepresentation of non-employed women at the bottom and top of the female wage distribution at age 42 in this cohort, suggests that the wife not working may have been a desirable option if the family could afford it. For this earliest cohort we study, accounting for selection into employment in wage comparisons reveals the hidden extent of gender inequalities in the labour market.

A major transition in women's employment and pay occurred after the 1970s, with the introduction of equal opportunities legislation and maternity leave. There was also increased scope for family planning as the contraceptive pill became more available. This historical shift weakens the argument that women's lower rates of pay, compared to men, are the natural and inevitable consequence of caring for children, and instead demonstrates the possibility for change in social and institutional arrangements. To assess and quantify the full impact of such changes on the position of women and men in work, it is important to use unbiased population-level indicators, as we have done. Our evidence is consistent with the view that improved labour market opportunities for women had an impact both on their pay and employment, and that we understate their impact if we do not account for selectivity biases in wage comparisons.

Acknowledgements

The funding for this work was provided via an Economic and Social Research Council (ESRC) studentship and the work was linked to the ESRC-funded Gender Equalities Network (GeNet). Thanks are due to everyone at the Centre for Longitudinal Studies and the MRC National Survey of Health and Development engaged in running the cohort studies and to cohort members themselves. We are grateful to Lorraine Dearden, Anna Vignoles, Sara Connolly, John Micklewright and Mark Killingsworth for useful comments.

References

Blau FD and Kahn LK. (2006) The U.S. gender pay gap in 1990s: slowing convergence. *Industrial and Labour Relations Review,* 60, 45-66.

Blossfeld H-P and Timm A. (2003) *Who marries whom? Educational systems as marriage markets in modern societies.* pp 9-18. European Studies in Population, Dordrecht, The Netherlands.

Blundell R, Gosling A, Ichimura H and Meghir C. (2007) Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica,* 75, 323-363.

Connolly S and Gregory M. (2009) The part-time pay penalty: earnings trajectories of British women. *Oxford Economic Papers*, 61, i76-i97.

Disney R and Webb S. (1991) Why are there so many long term sick in Britain? *The Economic Journal,* 3, 252-262.

Elliott J and Shepherd P. (2006) Cohort profile: 1970 British Birth Cohort (BCS70). *International Journal of Epidemiology,* 35, 836-843.

Ermisch J. (1988) Economic influences on birth rates. *National Institute Economic Review,* 126, 71-88.

Faggio G and Nickell S. (2003) The rise in inactivity among adult men. In R Dickens, P Gregg and J Wadsworth. eds. *The labour market under New Labour: The state of working Britain.* Pp 40-52. Palgrave Macmillan, London.

Ferri E, Bynner J and Wadsworth M. (2003) *Changing Britain, changing lives. Three generations at the turn of the century.* Institute of Education, London.

Gregg P, Gutierrez-Domenech M and Waldfogel J. (2007) The employment of married mothers in Great Britain, 1974-2000. *Economica,* 75, 842-864.

Gronau R. (1974) Wage comparisons – a selectivity bias. *The Journal of Political Economy,* 82, 1119-1143.

Harkness S. (2005) Pay inequality and gender. In S Delorenzi, J Reed, P Robinson. eds. *Maintaining Momentum: Promoting social mobility and life chances from early years to adulthood.* Institute for Public Policy Research, London.

Hawkes D and Plewis I. (2006) Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society,* Series A, 169, 479-492.

Heckman JJ. (1977) Sample selection bias as a specification error with an application to the estimation of labor supply functions. NBER Working Paper 172, National Bureau of Economic Research.

Heckman JJ. (1979) Sample selection bias as a specification error. *Econometrica,* 47, 153-162.

Joshi H. (1986) Participation in Paid Work: Evidence from the Women and Employment Survey. In R Blundell and I Walker. eds. *Unemployment, Search and Labour Supply.* Pp 217-212. Cambridge University Press, Cambridge.

Joshi H and Paci P. (1998) *Unequal pay for women and men: Evidence from the British Birth Cohort Studies.* The MIT Press, Cambridge, Ma.

Joshi H, Makepeace G and Dolton P. (2007) More or less unequal? Evidence on the pay of men and women from the British Birth Cohort Studies. *Gender, Work and Organization,* 14, 37-55.

Joshi H, Layard R and Owen S. (1985) Why are more women working in Britain? *Journal of Labor Economics,* 3, S147-S176.

Ketende S, McDonald J and Dex S. (2010) Non-response in the 1970 British Cohort Study (BCS70) from birth to 34 years. CLS Working Paper, Centre for Longitudinal Studies. http://www.cls.ioe.ac.uk/library (last accessed: 23 December 2010).

Kuh D and Wadsworth M. (1991) Childhood influences on adult male earnings in a longitudinal study. *The British Journal of Sociology,* 42, 537-555.

Kuh D, Head J, Hardy R and Wadsworth M. (1997) The influence of education and family background on women's earnings in midlife: evidence from a British national birth cohort study. *British Journal of Sociology of Education*, 18, 385-405.

Kuh D, Pierce M, Adams J, Deanfield J, Ekelund U, Friberg P, Ghosh AK, Harwood N, Hughes A, MacFarlane P, Mishra G, Pellerin D, Stephen AM, Wong A, Richards M and Hardy R (2011). Cohort Profile: Updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. *International Journal of Epidemiology*, 40, e1-9.

Leuven E and Sianesi B. (2003) PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing and covariate imbalance testing, version 3.1.5. Technical report.

Little RJA and Rubin DB. (1987) *Statistical Analysis with Missing Data.* Wiley, New York.

Makepeace G, Dolton P, Woods L, Joshi H and Galinda-Rueda F. (2003) From school to the labour market. In E Ferri, J Bynner and M Wadsworth. eds. *Changing Britain, changing lives. Three generations at the turn of the century.* Institute of Education, London.

Manning A. (1996) The Equal Pay Act as an experiment to test theories of the labour market. *Economica,* 63, 191-212.

Manning A and Petrongolo B. (2008) The part-time penalty for women in Britain. *The Economic Journal,* 118, F28-F51.

Manning A and Swaffield J. (2008) The gender pay gap in early-career wage growth. *The Economic Journal,* 118, 983-1024.

Manski CF. (1989) Anatomy of the selection problem. *The Journal of Human Resources,* 24, 343-360.

Neuburger J. (2010) Trends in the unequal pay of women and men across three British generations. PhD thesis, Institution of Education, University of London.

Olivetti C and Petrongolo B. (2008) Unequal pay or unequal employment? A cross-country analysis of gender gaps. *Journal of Labor Economics,* 26, 621-654.

Plewis I, Calderwood L, Hawkes D and Nathan G. (2004) National Child Development Survey and 1970 British Cohort Study technical report: Changes in the NCDS and BCS70 populations and samples over time. CLS Working Paper, Centre for Longitudinal Studies. http://www.cls.ioe.ac.uk/library.

Power C and Elliott J. (2006) Cohort profile: 1958 British birth cohort (National Child Development Survey). *International Journal of Epidemiology,* 35, 34-41.

Rosenbaum PR and Rubin DB. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika,* 70, 41-55.

Wadsworth MEJ, Butterworth S, Hardy R, Kuh D, Richards M, Langenberg C, Hilder WS and Connor M. (2003) The life course prospective design; an example of benefits and problems associated with study longevity. *Social Science and Medicine,* 57, 2193-5.

Wadsworth M, Kuh D, Richards M and Hardy R. (2006) Cohort profile: The 1946 national birth cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology,* 35, 49-94.

Wright RE and Ermisch JF. (1991) Gender Discrimination in the British Labour Market: A Reassessment. *Economic Journal,* 101, 508-422.

Zabalza A and Tzannatos Z. (1985) *Women and equal pay: The effects of legislation on female employment and wages in Britain.* Cambridge University Press, Cambridge.

Appendix - Additional Tables

Table A1. Variable definitions

Gross hourly wage	Derived gross wage per hour worked (including overtime) (£, 2000)
Part-time worker	Dummy = 1 if self-defined part-time worker (<30 hours a week)
Work experience	Years in paid employment up to time of survey since age 16 for the 1958 and 1970 cohorts and since age 25 for the 1946 cohort
Full-time experience	Years in full-time paid employment since age 16 (1958 and 1970 cohorts) or 25 (1946 cohort). Full-time work is self-defined
Part-time experience	Years in part-time paid employment since age 16 (1958 and 1970 cohorts) or 25 (1946 cohort). Part-time work is self-defined
Job tenure	Years working for current employer at time of survey
O-level or equivalent	Dummy = 1 if highest qualification = O-level or equivalent at time of survey
A-level or equivalent	Dummy = 1 if highest qualification = A-level or equivalent at time of survey
Diploma	Dummy = 1 if highest qualification = diploma from non-degree higher education at time of survey
Degree or higher	Dummy = 1 if highest qualification = Bachelors degree, equivalent or higher at time of survey
Maths score at age 11	Standardised score (z score) from maths test taken at age 10 (1970 cohort) or 11 (1946 and 1958 cohorts)
Missing maths score	Dummy = 1 if maths test not taken or score from test missing
London or SE	Dummy = 1 if living in London or the South East at time of survey.
Children in hhld	Dummy = 1 if own or other children living in household at time of survey
Young child	Dummy = 1 if child under 5 years of age living in household at time of survey
More than one child	Dummy = 1 if more than one child living in household at time of survey
Social class of first job	
I	Dummy=1 if first job in RG Class I
II	Dummy=1 if first job in RG Class II
III	Dummy=1 if first job in RG Class III
IV	Dummy=1 if first job in RG Class IV
V	Reference category (first job in RG Class V)
VI	Dummy=1 if first job in RG Class VI
Missing	Dummy=1 if information of occupation of 1st job missing

Table A1 (continued). Variable definitions (childhood information)

CM father in non-manual job (1946 cohort)	Dummy = 1 if cohort member (CM) born to father in non-manual occupation
I	Dummy=1 if first job in RG Class I
II	Dummy=1 if first job in RG Class II
III	Dummy=1 if first job in RG Class III
IV	Dummy=1 if first job in RG Class IV
V	Reference category (first job in RG Class V)
CM Mother's age:	
Youngest quartile	Reference category
Second quartile	Dummy = 1 if cohort member born to mother in second
Third quartile	Dummy = 1 if mother's age in third quartile of age distribution
Oldest quartile	Dummy = 1 if mother's age in top quartile of age distribution
Missing	Dummy = 1 if information on mother's age at birth missing
CM mother's education:	
Left before 16	Reference category
Left at 17	Dummy = 1 if mother of cohort member left school at age 17
Left at 18	Dummy = 1 if mother left school at age 18 or older
Missing	Dummy = 1 if information on mother's schooling missing
CM father's education:	
Left before 16	Reference category
Left at 17	Dummy = 1 if father of cohort member left school at age 17
Left at 18	Dummy = 1 if father left school at age 18 or older
Missing	Dummy = 1 if information on father's schooling missing
Number of siblings at age 16	
Only child	Dummy = 1 if cohort member had no siblings at age 16
One sibling	Dummy = 1 if one sibling at age 16
Two or three siblings	Dummy = 1 if two or three siblings at age 16
Four or more siblings	Reference category

Table A2. Difference in mean (age 10/11) maths scores by response at each survey

Birth cohort	Survey (Age)	Women	Men
1946	1972 (26)	+0.42 (0.32, 0.54)	+0.32 (0.22, 0.42)
1946	1977 (31)	+0.37 (0.27, 0.47)	+0.26 (0.16, 0.35)
1946	1989 (43)	+0.23 (0.14, 0.33)	+0.28 (0.19, 0.38)
1958	1981 (23)	+0.26 (0.21, 0.32)	+0.30 (0.25, 0.35)
1958	1991 (33)	+0.31 (0.27, 0.36)	+0.32 (0.27, 0.37)
1958	1999 (42)	+0.31 (0.27, 0.36)	+0.36 (0.32, 0.41)
1970	1996 (26)	+0.35 (0.30, 0.39)	+0.27 (0.23, 0.32)
1970	1999 (30)	+0.26 (0.21, 0.31)	+0.20 (0.15, 0.25)
1970	2003 (34)	+0.34 (0.29, 0.39)	+0.41 (0.36, 0.46)

Notes: The figures shown are mean standardised maths scores (standard deviation = 1) for non-respondents deducted from mean scores for respondents. 95 % confidence intervals are shown in brackets. Non-respondents include refusals and non-contacts, but exclude emigrants and those who had died. Missing maths scores were estimated using previous childhood variables.

Table A3. Self-reported main economic activity, by gender, age at survey and cohort

	1946 cohort			1958 cohort			1970 cohort		
	26	31	43	23	33	42	26	30	34
Women									
Full-time employee	31	21	37	56	33	40	62	48	40
Part-time employee	13	29	34	6	29	31	12	21	29
FT self-employed	1	3	6	-	4	4	3	3	3
PT self-employed	1	2	2	-	3	3	1	2	3
Housewife / carer	51	45	11	24	27	13	14	20	19
Unemployed	1	1	4	7	2	2	2	2	2
Other not in work	-	-	3	-	1	1	2	1	1
Full-time student	1	1	1	2	1	1	2	1	2
Other	-	-	-	2	1	1	2	1	2
Sample size	1,852	1,649	1,618	6,256	5,785	5,777	4,835	5,766	5,025
Men									
Full-time employee	84	83	70	75	74	71	75	77	77
Part-time employee	-	-	1	1	1	1	2	1	1
FT self-employed	10	12	21	6	16	17	10	11	14
PT self-employed	-	-	-	-	-	1	1	1	1
Housewife / carer	-	-	-	-	-	1	-	1	1
Unemployed	3	4	2	12	6	3	7	5	3
Other not in work	1	-	4	-	2	5	2	3	3
Full-time student	1	1	-	3	-	-	4	1	1
Other	-	-	-	2	1	1	1	1	1
Sample size	1,897	1,661	1,607	6,249	5,582	5,605	4,063	5,436	4,609

Notes: Percentages are rounded to the nearest integer and are not reported when less than half a percent of the sample fell into the specified category. Percentages for the 1946 cohort are weighted to given population estimates.

Table A4. Regression parameters (std. errors) from probit models used to estimate propensity scores for women, DV = 1 for non-employee

	1946 cohort			1958 cohort			1970 cohort		
Age at survey	*26*	*31*	*43*	*23*	*33*	*42*	*26*	*30*	*34*
1946 cohort weight	-0.09 (0.09)	+0.21 (0.10)	-0.21 (0.13)	-	-	-	-	-	-
CM mother's age									
Youngest quartile	reference	reference	reference	reference	reference	reference	reference	reference	reference
2nd quartile	+0.06 (0.19)	+0.15 (0.11)	+0.05 (0.16)	-0.06 (0.07)	+0.01 (0.06)	-0.07 (0.05)	-0.02 (0.10)	+0.09 (0.06)	+0.11 (0.06)
3rd quartile	+0.14 (0.12)	+0.03 (0.11)	-0.01 (0.15)	-0.09 (0.07)	+0.02 (0.06)	-0.01 (0.06)	+0.03 (0.10)	+0.09 (0.06)	+0.11 (0.06)
Oldest quartile	+0.13 (0.13)	+0.04 (0.12)	+0.06 (0.16)	-0.02 (0.07)	+0.04 (0.06)	-0.08 (0.06)	+0.17 (0.09)	+0.03 (0.07)	+0.12 (0.07)
Missing age	-0.22 (0.18)	+0.04 (0.20)	0.00 (0.26)	-0.39 (0.12)	-0.03 (0.11)	-0.22 (0.11)	-0.50 (0.41)	+0.56 (0.07)	-0.29 (0.46)
CM father's social class									
I	+0.22 (0.23)	-0.21 (0.22)	-0.18 (0.30)	+0.08 (0.13)	-0.11 (0.11)	+0.11 (0.10)	-0.39 (0.16)	+0.12 (0.11)	+0.13 (0.11)
II	+0.25 (0.14)	0.00 (0.14)	-0.22 (0.19)	-0.06 (0.09)	-0.07 (0.08)	+0.09 (0.07)	-0.38 (0.13)	+0.07 (0.08)	+0.01 (0.08)
III	+0.34 (0.13)	-0.20 (0.13)	+0.06 (0.18)	-0.05 (0.10)	-0.03 (0.08)	-0.05 (0.08)	-0.23 (0.13)	+0.12 (0.10)	+0.07 (0.09)
IV	0.00 (0.11)	-0.05 (0.11)	-0.18 (0.15)	+0.02 (0.06)	-0.19 (0.06)	-0.08 (0.05)	-0.28 (0.10)	+0.11 (0.07)	0.00 (0.06)
V & VI	reference	reference	reference	reference	reference	reference	reference	reference	reference
Missing social class	0.00 (0.18)	-0.57 (0.20)	+0.07 (0.25)	+0.02 (0.09)	-0.16 (0.07)	+0.05 (0.08)	-0.31 (0.26)	-0.08 (0.18)	-0.02 (0.17)
CM mother's schooling									
Left at 16 or younger	reference	reference	reference	reference	reference	reference	reference	reference	reference
Left at 17	-0.08 (0.21)	-0.10 (0.21)	+0.09 (0.30)	-0.01 (0.13)	+0.04 (0.12)	-0.10 (0.13)	-0.06 (0.10)	-0.17 (0.09)	+0.05 (0.08)
Left at 18	0.00 (0.28)	+0.12 (0.26)	+0.57 (0.31)	-0.08 (0.13)	0.00 (0.11)	+0.01 (0.10)	0.00 (0.11)	-0.02 (0.10)	-0.10 (0.09)
Schooling missing	+0.51 (0.29)	+0.72 (0.37)	+0.19 (0.40)	0.00 (0.18)	-0.01 (0.17)	-0.10 (0.18)	+0.68 (0.40)	+0.54 (0.37)	+0.16 (0.42)
CM father's schooling									
Left at 16 or younger	reference	reference	reference	reference	reference	reference	reference	reference	reference
Left at 17	+0.30 (0.20)	0.00 (0.19)	-0.02 (0.27)	-0.16 (0.14)	+0.09 (0.11)	-0.01 (0.14)	+0.04 (0.11)	-0.11 (0.10)	+0.04 (0.09)
Left at 18	-0.29 (0.23)	-0.08 (0.20)	0.00 (0.21)	+0.13 (0.11)	+0.01 (0.09)	0.00 (0.09)	+0.17 (0.10)	+0.10 (0.09)	+0.04 (0.09)
Schooling missing	-0.39 (0.29)	-0.76 (0.36)	+0.15 (0.38)	-0.12 (0.14)	-0.26 (0.12)	+0.12 (0.12)	+0.04 (0.28)	+0.26 (0.17)	-0.03 (0.17)
CM siblings, at age 16									
Only child	-0.12 (0.18)	+0.07 (0.15)	+0.07 (0.22)	-0.15 (0.13)	+0.04 (0.11)	-0.04 (0.10)	-0.25 (0.16)	-0.14 (0.13)	+0.12 (0.13)
One sibling	+0.05 (0.18)	+0.07 (0.14)	+0.10 (0.18)	+0.02 (0.13)	+0.12 (0.11)	+0.09 (0.08)	-0.22 (0.14)	-0.14 (0.11)	+0.15 (0.11)

Table A4 (continued). Regression parameters (std. errors) from probit models used to estimate propensity scores for women

	1946 cohort			1958 cohort			1970 cohort		
Age at survey	*26*	*31*	*43*	*23*	*33*	*42*	*26*	*30*	*34*
Two or three siblings	+0.08 (0.12)	+0.05 (0.12)	+0.09 (0.16)	+0.02 (0.07)	+0.12 (0.07)	+0.07 (0.06)	-0.22 (0.14)	-0.23 (0.10)	+0.10 (0.10)
Four or more siblings	reference	reference	reference	reference	reference	reference	reference	reference	reference
Siblings missing	-0.08 (0.23)	+0.32 (0.24)	-0.39 (0.28)	+0.07 (0.18)	+0.32 (0.17)	+0.07 (0.18)	-0.18 (0.15)	-0.10 (0.12)	+0.19 (0.13)
Maths score, age 10/11	-0.12 (0.05)	-0.03 (0.05)	-0.03 (0.07)	-0.09 (0.03)	-0.06 (0.03)	-0.02 (0.03)	-0.01 (0.03)	-0.06 (0.03)	-0.03 (0.03)
Missing score	+0.31 (0.16)	+0.05 (0.17)	+0.50 (0.22)	-0.02 (0.07)	-0.10 (0.06)	-0.01 (0.06)	-0.05 (0.07)	+0.02 (0.05)	+0.04 (0.05)
Highest qualification									
No qualifications	reference	reference	reference	reference	reference	reference	reference	reference	reference
O-level or equivalent	+0.11 (0.11)	+0.05 (0.11)	-0.15 (0.15)	-0.31 (0.06)	-0.01 (0.06)	+0.01 (0.06)	-0.18 (0.07)	-0.12 (0.06)	-0.10 (0.07)
A-level or equivalent	-0.18 (0.14)	+0.22 (0.14)	-0.12 (0.18)	-0.76 (0.10)	-0.11 (0.08)	-0.09 (0.11)	-0.21 (0.11)	+0.04 (0.07)	-0.12 (0.08)
Diploma	+0.23 (0.17)	+0.47 (0.16)	-0.45 (0.24)	-0.81 (0.10)	-0.29 (0.08)	-0.20 (0.08)	-0.45 (0.12)	-0.28 (0.28)	-0.46 (0.08)
Degree or higher	+0.05 (0.22)	+0.19 (0.23)	-0.04 (0.22)	-1.24 (0.13)	-0.73 (0.10)	-0.31 (0.09)	-0.09 (0.11)	-0.91 (0.09)	-0.84 (0.09)
Years in full-time work	-0.07 (0.01)	-0.25 (0.02)	-0.17 (0.01)	-0.34 (0.02)	-0.15 (0.01)	-0.11 (0.00)	-0.08 (0.01)	-0.18 (0.01)	-0.15 (0.01)
Years in part-time work	-	-0.20 (0.03)	-0.22 (0.02)	-0.86 (0.05)	-0.26 (0.01)	-0.16 (0.01)	-0.14 (0.02)	-0.26 (0.01)	-0.20 (0.01)
Children in hhld	+0.35 (0.17)	+0.98 (0.13)	-0.14 (0.16)	+0.58 (0.16)	+0.22 (0.08)	-0.09 (0.07)	+1.02 (0.11)	+0.41 (0.07)	+0.17 (0.08)
Children under five	+1.29 (0.15)	-	+1.31 (0.25)	+1.36 (0.15)	+0.92 (0.05)	+0.87 (0.07)	+0.31 (0.11)	+0.75 (0.07)	+0.77 (0.06)
More than 1 child	-0.15 (0.10)	-	-0.33 (0.14)	-0.19 (0.09)	-0.12 (0.06)	-0.18 (0.06)	+0.30 (0.10)	+0.01 (0.06)	+0.11 (0.06)
Living in London/SE	-0.10 (0.08)	+0.27 (0.09)	-0.02 (0.11)	-0.06 (0.13)	+0.08 (0.05)	+0.02 (0.05)	-	+0.12 (0.05)	+0.14 (0.05)
Social class of 1st job									
I	-	-	-	-0.96 (0.26)	+0.57 (0.17)	+0.12 (0.21)	-0.57 (0.21)	-0.88 (0.21)	-0.44 (0.18)
II	-	-	-	-0.39 (0.10)	-0.04 (0.07)	-0.21 (0.09)	-0.45 (0.08)	-0.37 (0.08)	-0.20 (0.08)
III	-	-	-	+0.08 (0.10)	+0.01 (0.05)	+0.02 (0.06)	-0.19 (0.07)	-0.18 (0.06)	-0.08 (0.06)
IV	-	-	-	+0.21 (0.07)	+0.12 (0.20)	-0.21 (0.08)	+0.02 (0.09)	-0.04 (0.08)	-0.11 (0.09)
V	-	-	-	reference	reference	reference	reference	reference	reference
VI	-	-	-	+0.14 (0.26)	-0.07 (0.23)	-0.05 (0.23)	+0.01 (0.18)	-0.32 (0.16)	-0.43 (0.17)
SC missing	-	-	-	+1.61 (0.22)	+0.18 (0.11)	-0.50 (0.09)	+0.18 (0.12)	+0.47 (0.16)	-0.15 (0.11)
Constant term	-0.61 (0.18)	-0.29 (0.20)	+1.41 (0.28)	+1.03 (0.14)	+0.88 (0.12)	+1.34 (0.12)	-0.55 (0.19)	+0.77 (0.15)	+0.79 (0.17)
Sample size	1,660	1,340	1,079	5,626	5,301	5,339	3,684	5,473	4,714

Table A5. Regression parameters (std. errors) from probit models used to estimate propensity scores for men, DV = 1 if non-employee

	1946 cohort			1958 cohort			1970 cohort		
Age at survey	*26*	*31*	*43*	*23*	*33*	*42*	*26*	*30*	*34*
1946 cohort weight	0.00 (0.18)	+0.10 (0.09)	-0.37 (0.23)	-	-	-	-	-	-
CM mother's age									
Youngest quartile	reference	reference	reference	reference	reference	reference	reference	reference	reference
2nd quartile	+0.03 (0.21)	-0.12 (0.21)	+0.18 (0.29)	-0.04 (0.07)	-0.13 (0.08)	+0.12 (0.08)	+0.07 (0.07)	-0.03 (0.06)	+0.06 (0.11)
3rd quartile	+0.28 (0.19)	+0.06 (0.19)	+0.17 (0.29)	-0.18 (0.10)	-0.03 (0.08)	-0.11 (0.09)	+0.01 (0.08)	-0.01 (0.07)	+0.03 (0.11)
Oldest quartile	-0.12 (0.23)	-0.20 (0.21)	+0.48 (0.28)	-0.17 (0.10)	-0.06 (0.08)	-0.08 (0.09)	+0.18 (0.10)	-0.02 (0.07)	-0.06 (0.11)
Missing age	-0.48 (0.34)	+0.37 (0.31)	+0.46 (0.43)	-0.33 (0.12)	-0.04 (0.14)	-0.06 (0.15)	-0.50 (0.41)	-0.14 (0.31)	+0.33 (0.32)
CM father's social class									
I	-0.10 (0.37)	+0.31 (0.38)	-0.71 (0.70)	+0.19 (0.13)	-0.12 (0.11)	-0.23 (0.19)	-0.39 (0.16)	-0.01 (0.11)	+0.22 (0.11)
II	-0.20 (0.31)	+0.07 (0.26)	-0.03 (0.33)	+0.18 (0.10)	-0.13 (0.11)	-0.19 (0.07)	-0.38 (0.13)	+0.08 (0.08)	+0.22 (0.08)
III	-0.21 (0.28)	-0.25 (0.28)	+0.13 (0.32)	+0.17 (0.11)	-0.08 (0.12)	-0.30 (0.08)	-0.23 (0.15)	+0.05 (0.10)	-0.04 (0.10)
IV	+0.45 (0.19)	+0.19 (0.18)	-0.05 (0.11)	+0.05 (0.07)	+0.01 (0.08)	+0.02 (0.05)	-0.28 (0.10)	-0.12 (0.06)	+0.02 (0.07)
V & VI	reference	reference	reference	reference	reference	reference	reference	reference	reference
Missing social class	-0.14 (0.35)	-0.02 (0.37)	-0.18 (0.45)	-0.09 (0.09)	+0.11 (0.09)	+0.02 (0.07)	-0.31 (0.27)	-0.05 (0.16)	-0.03 (0.17)
CM mother's schooling									
Left at 16 or younger	reference	reference	reference	reference	reference	reference	reference	reference	reference
Left at 17	-0.22 (0.22)	-0.10 (0.38)	-0.35 (0.53)	-0.16 (0.15)	-0.03 (0.18)	+0.14 (0.11)	+0.04 (0.10)	+0.09 (0.30)	0.00 (0.09)
Left at 18	-0.26 (0.56)	+0.20 (0.45)	+0.27 (0.63)	-0.03 (0.13)	+0.03 (0.18)	+0.13 (0.11)	+0.05 (0.11)	+0.02 (0.09)	-0.08 (0.10)
Schooling missing	+0.23 (0.65)	+0.05 (0.57)	+0.70 (0.70)	-0.02 (0.19)	+0.28 (0.22)	+0.12 (0.15)	+0.08 (0.37)	+0.09 (0.30)	-0.21 (0.32)
CM father's schooling									
Left at 16 or younger	reference	reference	reference	reference	reference	reference	reference	reference	reference
Left at 17	+0.33 (0.32)	-0.13 (0.38)	-0.35 (0.53)	-0.05 (0.16)	+0.16 (0.19)	+0.14 (0.12)	+0.13 (0.11)	-0.13 (0.10)	-0.08 (0.10)
Left at 18	+0.32 (0.34)	+0.42 (0.34)	+0.26 (0.63)	+0.13 (0.12)	-0.16 (0.18)	-0.05 (0.09)	+0.13 (0.10)	-0.25 (0.10)	-0.06 (0.09)
Schooling missing	+0.03 (0.65)	-0.70 (0.56)	-0.59 (0.70)	+0.30 (0.14)	-0.17 (0.16)	+0.09 (0.12)	+0.20 (0.20)	+0.27 (0.16)	-0.02 (0.17)
CM siblings, at age 16									
Only child	-0.86 (0.36)	-0.33 (0.31)	-2.06 (1.69)	-0.11 (0.13)	-0.02 (0.11)	+0.10 (0.11)	-0.33 (0.17)	+0.12 (0.14)	+0.22 (0.14)
One sibling	-0.22 (0.22)	-0.02 (0.22)	+0.86 (0.36)	-0.12 (0.13)	0.00 (0.08)	+0.10 (0.08)	-0.32 (0.15)	0.00 (0.12)	+0.10 (0.12)

Table A5 (continued). Regression parameters (std. errors in brackets) from probit models used to estimate propensity scores for men

	1946 cohort			1958 cohort			1970 cohort		
Age at survey	*26*	*31*	*43*	*23*	*33*	*42*	*26*	*30*	*34*
Two or three siblings	-0.25 (0.20)	-0.07 (0.21)	+0.67 (0.36)	-0.04 (0.08)	+0.04 (0.06)	+0.03 (0.06)	-0.27 (0.14)	+0.04 (0.10)	+0.16 (0.11)
Four or more	reference	reference	reference	reference	reference	reference	reference	reference	reference
Siblings missing	+0.11 (0.36)	+0.51 (0.57)	+1.16 (0.55)	-0.32 (0.17)	-0.05 (0.22)	-0.02 (0.22)	-0.03 (0.18)	+0.04 (0.15)	+0.15 (0.20)
Maths score, age 10/11	-0.02 (0.08)	-0.31 (0.09)	+0.11 (0.11)	-0.01 (0.29)	-0.04 (0.03)	-0.05 (0.04)	-0.02 (0.04)	-0.02 (0.03)	-0.09 (0.04)
Missing score	-0.28 (0.29)	-1.15 (0.48)	-0.44 (0.54)	+0.01 (0.07)	-0.02 (0.05)	+0.04 (0.09)	-0.07 (0.08)	+0.11 (0.05)	+0.05 (0.09)
Highest qualification									
No qualifications	reference	reference	reference	reference	reference	reference	reference	reference	reference
O-level or equivalent	+0.05 (0.23)	+0.27 (0.22)	-0.25 (0.30)	-0.31 (0.07)	-0.13 (0.08)	+0.03 (0.08)	-0.21 (0.08)	-0.08 (0.08)	+0.15 (0.11)
A-level or equivalent	-0.07 (0.23)	-0.14 (0.26)	-0.51 (0.32)	-0.65 (0.08)	-0.42 (0.10)	-0.08 (0.10)	-0.35 (0.12)	-0.25 (0.09)	+0.05 (0.11)
Diploma	-0.29 (0.30)	-0.26 (0.31)	-0.10 (0.31)	-0.80 (0.11)	-0.52 (0.08)	-0.24 (0.11)	-0.30 (0.13)	-0.32 (0.11)	-0.02 (0.12)
Degree or higher	-0.29 (0.30)	-0.07 (0.32)	-0.69 (0.37)	-1.60 (0.12)	-1.03 (0.12)	-0.93 (0.13)	-0.02 (0.11)	-1.08 (0.12)	-0.45 (0.13)
Years in full-time work	-0.28 (0.02)	-0.41 (0.06)	-0.41 (0.04)	-0.45 (0.02)	-0.17 (0.01)	-0.14 (0.01)	-0.04 (0.01)	-0.21 (0.04)	-0.17 (0.01)
Years in part-time work	-	+0.13 (0.15)	-0.48 (0.12)	-0.65 (0.14)	-0.15 (0.04)	-0.19 (0.03)	-0.05 (0.05)	-0.24 (0.04)	-0.08 (0.03)
Children in hhld	+0.25 (0.40)	-	-0.51 (0.21)	+0.24 (0.24)	+0.20 (0.10)	-0.23 (0.09)	+0.24 (0.18)	-0.03 (0.12)	-0.23 (0.13)
Children under five	-0.37 (0.40)	-	-1.14 (0.54)	-0.17 (0.25)	-0.11 (0.08)	+0.15 (0.09)	-0.66 (0.19)	-0.10 (0.12)	-0.14 (0.11)
More than 1 child	+0.15 (0.23)	-	+0.12 (0.29)	+0.14 (0.13)	+0.10 (0.09)	-0.08 (0.09)	+0.19 (0.17)	+0.18 (0.10)	+0.27 (0.12)
Living in London/SE	-0.32 (0.18)	+0.02 (0.16)	-0.20 (0.23)	+0.16 (0.08)	+0.02 (0.08)	-0.13 (0.07)	-	-0.06 (0.07)	+0.07 (0.08)
Social class of 1st job									
I	-	-	-	-0.29 (0.10)	-0.36 (0.21)	-0.11 (0.21)	-0.59 (0.16)	-0.81 (0.21)	-0.48 (0.24)
II	-	-	-	-0.14 (0.06)	-0.25 (0.08)	-0.13 (0.14)	-0.57 (0.10)	-0.61 (0.11)	-0.19 (0.13)
III	-	-	-	-0.17 (0.06)	-0.15 (0.06)	-0.18 (0.11)	-0.36 (0.10)	-0.28 (0.10)	-0.09 (0.12)
IV	-	-	-	+0.09 (0.06)	-0.09 (0.05)	-0.07 (0.09)	-0.24 (0.09)	-0.04 (0.08)	+0.23 (0.10)
V	-	-	-	reference	reference	reference	reference	reference	reference
VI	-	-	-	+0.14 (0.10)	-0.15 (0.12)	0.00 (0.12)	+0.44 (0.14)	+0.19 (0.11)	+0.40 (0.14)
SC missing	-	-	-	+0.48 (0.12)	-0.26 (0.12)	-0.68 (0.12)	+0.13 (0.12)	+0.13 (0.13)	+0.13 (0.13)
Constant term	+1.32 (0.33)	+0.53 (0.38)	+5.10 (0.73)	+1.89 (0.14)	+1.04 (0.12)	+1.87 (0.17)	-0.19 (0.21)	+1.31 (0.19)	+0.45 (0.16)
Sample size	1,602	1,272	996	5,219	4,580	4,591	3,034	4,752	3,928

Table A6. Difference in mean maths scores and model fit before and after matching

Survey	**Women**		**Men**	
(Age)	Before matching	After matching	Before matching	After matching
(1) Mean maths scores				
1972 (26)	+0.26 (0.17, 0.35)	-0.04 (-0.07, 0.15)	+0.14 (-0.11, 0.39)	-0.14 (-0.52 0.24)
1977 (31)	+0.06 (-0.04, 0.16)	+0.01 (-0.11, 0.13)	+0.55 (0.29, 0.82)	-0.02 (-0.43, 0.38)
1989 (43)	+0.20 (0.05, 0.35)	-0.12 (-0.34, 0.10)	+0.19 (-0.14, 0.53)	+0.29 (-0.25, 0.83)
1981 (23)	+0.42 (0.36, 0.47)	-0.09 (-0.16, -0.02)	0.00 (-0.08, 0.08)	+0.05 (-0.07, 0.18)
1991 (33)	+0.22 (0.17, 0.27)	-0.03 (-0.11, 0.04)	+0.48 (0.39, 0.58)	+0.02 (-0.11, 0.15)
1999 (42)	+0.21 (0.15, 0.27)	-0.04 (-0.14, 0.05)	+0.48 (0.39, 0.58)	-0.11 (-0.24, 0.02)
1996 (26)	+0.23 (0.16, 0.30)	-0.05 (-0.16, 0.06)	+0.11 (0.01, 0.22)	-0.04 (-0.19, 0.11)
1999 (30)	+0.31 (0.25, 0.36)	+0.04 (-0.04, 0.12)	+0.25 (0.15, 0.35)	+0.13 (-0.01, 0.28)
2003 (34)	+0.19 (0.13, 0.26)	+0.04 (-0.05, 0.12)	+0.31 (0.20, 0.43)	-0.04 (-0.21, 0.13)
(2) Pseudo R2 from probit models				
1972 (26)	0.62	0.12	0.49	0.14
1977 (31)	0.25	0.03	0.23	0.10
1989 (43)	0.34	0.14	0.55	0.17
1981 (23)	0.52	0.06	0.29	0.04
1991 (33)	0.28	0.02	0.28	0.04
1999 (42)	0.27	0.02	0.33	0.04
1996 (26)	0.26	0.02	0.08	0.02
1999 (30)	0.35	0.02	0.31	0.03
2003 (34)	0.29	0.01	0.31	0.05
(3) Likelihood-Ratio test (p-value) of significance of variables				
1972 (26)	1233.4 (0.00)	152.2 (0.00)	332.2 (0.00)	24.0 (0.99)
1977 (31)	457.4 (0.00)	57.5 (0.00)	115.9 (0.00)	16.3 (0.95)
1989 (43)	355.3 (0.00)	76.5 (0.00)	270.8 (0.00)	17.4 (0.96)
1981 (23)	3824.2 (0.00)	296.5 (0.00)	1337.5 (0.00)	81.1 (0.00)
1991 (33)	1893.6 (0.00)	100.1 (0.00)	910.9 (0.00)	56.7 (0.02)
1999 (42)	1518.0 (0.00)	64.8 (0.00)	1066.0 (0.00)	53.1 (0.04)
1996 (26)	1012.7 (0.00)	45.4 (0.14)	178.5 (0.00)	22.1 (0.97)
1999 (30)	2250.0 (0.00)	92.5 (0.00)	1010.5 (0.00)	30.6 (0.76)
2003 (34)	1601.6 (0.00)	41.6 (0.28)	703.1 (0.00)	40.8 (0.31)

Notes: The figures shown are (1) mean standardised maths scores (95% confidence intervals) for non-employees deducted from mean scores for employees. (2) The pseudo R2 from a probit model regressing the propensity score on all the variables included in the model for the unmatched and matched (weighted) samples. (3) The likelihood ratio tests of the joint insignificance of these covariates before and after matching. These statistics were estimated using the pstest commands in Stata (Leuven and Sianesi 2003).

Table A7. Sample sizes for observed and imputed wages, by gender, age and cohort

	1946 cohort			1958 cohort			1970 cohort		
	26	31	43	23	33	42	26	30	34
Women									
Observed wage	739	621	1,129	3,657	3,185	4,000	3,237	3,903	3,005
Employed, missing wage	74	124	32	230	406	134	249	83	433
Not employed	884	679	193	1,901	1,761	1,205	811	1,453	1,277
No common support	2	5	16	132	6	5	1	52	3
Missing key covariates*	97	114	110	223	40	4	356	11	7
Sample size	1,796	1,543	1,480	6,143	5,398	5,348	4,654	5,502	4,725
Men									
Observed wage	1,463	1,181	1,105	4,375	3,754	3,958	2,779	4,183	3,291
Employed, missing wage	120	156	25	367	381	121	256	89	319
Not employed	68	60	38	788	517	509	387	444	321
No common support	20	2	29	69	12	7	0	81	11
Missing key covariates*	97	114	110	223	40	4	356	11	7
Sample size	1,715	1,441	1,240	5,837	4,696	4,607	3,652	4,824	3,950

**Missing wages were not imputed if items were missing for either highest qualification, employment experience or, for women only, number and ages of children.*

Table A8. Median observed and imputed potential wages (£, 2000 prices)

	1946 cohort			1958 cohort			1970 cohort		
	26	31	43	23	33	42	26	30	34
Women									
Observed wage	4.06	4.02	5.07	4.71	6.08	6.39	6.33	6.96	7.68
Employed, missing wage	4.17	4.42	5.07	4.63	5.62	5.85	5.94	6.83	7.12
Not employed	3.33	3.79	5.00	3.66	4.71	5.23	5.69	5.41	5.62
Men									
Observed wage	5.95	6.29	8.50	5.61	8.79	9.51	7.06	8.05	9.54
Employed, missing wage	6.14	6.72	9.36	5.65	7.75	10.43	6.90	7.06	9.45
Not employed	5.77	5.76	6.62	5.28	6.98	6.73	6.90	6.48	7.44

Notes: Medians for each 1946 cohort sample are weighted to give population estimates.

Table A9. % imputed wages below 25th, 50th and 75th percentile of employee wages

	Women			Men		
(Age)	25th	50th	75th	25th	50th	75th
1972 (26)	38.1%	70.4%	82.9%	36.1%	57.4%	69.0%
1977 (31)	37.3%	59.8%	77.8%	41.3%	60.0%	79.3%
1989 (43)	35.7%	50.1%	68.9%	34.7%	62.2%	78.6%
1981 (23)	54.1%	76.3%	89.1%	31.6%	58.1%	83.8%
1991 (33)	37.7%	67.5%	85.2%	44.3%	71.0%	87.6%
1999 (42)	40.5%	66.4%	82.7%	52.8%	74.5%	90.2%
1996 (26)	35.6%	59.2%	76.6%	26.8%	52.7%	78.6%
1999 (30)	46.3%	67.2%	83.0%	43.2%	63.2%	84.7%
2003 (34)	47.8%	69.8%	86.7%	46.1%	67.6%	85.4%

Notes: Imputed wages for non-employees. Weights used for 1946 cohort samples.

Table A10. % imputed wages below the median (1) nearest-neighbour (2) longitudinal

Survey		**Women**			**Men**	
(Age)	(1)	(2)	Sample (N)	(1)	(2)	Sample (N)
1972 (26)	71.4%	62.6%	560	54.9%	72.0%	37
1977 (31)	57.9%	52.6%	466	55.3%	65.8%	45
1989 (43)	50.0%	52.3%	97	63.9%	61.4%	32
1981 (23)	77.4%	70.6%	1,169	58.1%	56.5%	432
1991 (33)	64.4%	59.0%	1,194	65.9%	66.2%	293
1999 (42)	60.8%	53.6%	646	69.6%	67.8%	283
1996 (26)	57.5%	71.2%	472	51.3%	62.5%	269
1999 (30)	64.4%	65.1%	674	58.2%	68.2%	170
2003 (34)	64.4%	55.5%	735	58.1%	68.4%	136

Notes: The samples include non-employees at each survey with both (1) an imputed wage from propensity score matching and (2) an observed wage from another survey.
For this reason, percentages falling below the median are slightly different to those for the complete non-employed samples shown in Table A9. Weights used for 1946 cohort.