THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2008/27

Measuring and testing for gender discrimination in professions: the case of English family doctors

By

Hugh Gravelle and Arne Risa Hole, University of York

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Measuring and testing for gender discrimination in professions: the case of English family doctors

Hugh Gravelle* **Arne Risa Hole***

Abstract

In 2004 the income of female GPs was 70%, and their wages (income per hour) were 91%, of those of male GPs.

We compare estimates of gender discrimination from Oaxaca decompositions using models of wages (income/hours), OLS and 2SLS models of income, and propensity score matching. We propose a new direct test for within workplace gender discrimination based on a comparison of the differences in income of female and male GPs in practices in which all GPs are of the same gender with the differences in male and female income in mixed gender practices.

We find that the coefficients on log hours in the log income models are positive but significantly less than 1, so that log wage models are misspecified. Discrimination, as measured by the unexplained difference in mean log income varied between 21% to 28%. However, our direct tests could not reject the null hypothesis of no within workplace gender discrimination.

JEL Nos: J16, J44, J71, I11
Keywords: Gender discrimination. Professions. Family doctors. Propensity score matching.

* National Primary Care Research and Development Centre, Centre for Health Economics, University of York; emails: hg8@york.ac.uk; ah522@york.ac.uk. NPCRDC receives core funding from the Department of Health. The views expressed are those of the authors and not necessarily those of the funders. The NPCRDC General Practitioner Worklife Survey was undertaken by Diane Whalley, Chris Bojke, Hugh Gravelle and Bonnie Sibbald. We are grateful to Mark Dusheiko, Paul Chambers-Dixon, and Andrew Wagner for providing additional data; and to members of the York Health, Econometrics and Data Group, Steve Morris, Karen Mumford and Jo Swaffield for comments.

1 Introduction

In this paper we analyse the substantial differences in income between male and female general practitioners (GPs) in the English National Health Service. Our study is of interest for a number of reasons. First, it is the first study of income differentials in this group of workers who account for the majority of patient contacts with the NHS and who act as gatekeepers to NHS inpatient and outpatient care. Second, the effect of investment in human capital on income may depend, not just on gender, but on other worker characteristics. By restricting attention to groups of workers with similar human capital, as in recent studies of lawyers, academics and US physicians, it is easier to isolate the effect of gender. Indeed, since we examine a group of doctors within the same speciality in a national health system which imposes uniform qualification requirements, we have a group with more homogenous human capital than in other single occupation studies. Third, we consider a set of methodological issues (the choice between wages and income as the measure of pay, the endogeneity of hours of work, parametric versus non-parametric estimates of gender pay differences) and their implications for the estimated discrimination measures. We examine the robustness of estimates of discrimination between male and female GPs to alternative ways of dealing with these methodological issues.

We also suggest a new method of testing directly for pay discrimination. English family doctors typically work in small practices or partnerships of 3 to 4 GPs. The income of a GP depends on the total income of their practice and the within practice income sharing arrangements. GPs undertake a range of activities with different effects on practice income and patient health. Within practice discrimination can take three forms. Female GPs may be assigned a mix of activities which generate less income. The practice income sharing formula may give lower rewards to activities which women prefer or have a comparative advantage in. Or women may get a smaller reward for any given mix of activities. Our test for within practice discrimination rests on the fact that none of these forms of discrimination can arise in practices where all the GPs have the same gender. We compare the incomes of GPs in all male and all female practices to estimate the difference in income which reflects gender productivity or preference differences. The unexplained difference between male and female GP incomes in mixed gender practices, net of the difference between those in single gender practices, is then an estimate of the extent of within practice gender discrimination.

Section 2 discusses methodological questions in the measurement of gender pay discrimination. Section 3 describes the data and presents estimates based on alternative measures in section 4. Section 5 suggests a new way of testing whether differences are due to discrimination and applies it to English GPs.

The next part of the introduction summarises the relevant literature on gender differences in pay in professional occupations including medicine.

1.1 Gender differences in professional pay: related literature

There are gender differences in pay even for individuals with considerable investments in human capital working in the same profession (Wood, Corcoran, and Courant,1993; Dolton, O'Neill and Sweetman, 1996; Blackaby, Booth, and Frank, 2005; Noonan, Corcoran and Courant, 2005; McNabb and Wass, 2006). In medicine marked gender differences in income and hourly pay are reported in many countries (Robinson, 1998; Gupta et al, 2003).

In US medicine Kehrer (1976) found a 30% overall difference and a 22% unexplained difference in hourly wages in 1973. Langwell (1982) found similar (19%) unexplained difference, though a smaller (22%) overall difference, in wages for 1978 Ohsfeldt and Culler (1986) criticised these two studies for failing to correct for retransformation bias after estimating log wage models. They found that in 1982-3 the difference in hourly wage was 30%. Using an improved specification of the log wage model and allowing for retransformation bias, they estimate that 13% of the difference was unexplained.

Baker (1996) suggested that gender pay differentials had disappeared amongst young (under 45) US physicians in 1991. Although there was a 40% difference in income, the hourly pay difference was 14%. After conditioning on speciality and practice setting there was no significant difference in hourly pay. In primary care female physicians had 13% higher hourly wages. Bashaw and Heywood (2002) show that the reason why income differences are much larger than wage differences is that the hourly wage declines with hours worked and women work shorter hours. Using the same data as Baker (1996) they find a 17% unexplained difference in income. Ash et al (2004) find that female US medical academics are less likely to hold higher rank and have lower pay given their rank. Their pay is 13% less than if they were rewarded for their characteristics at the same rate as men.

2 Measuring gender discrimination

2.1 Regression based methods

The standard method for analysing the difference in pay between male and female workers is to estimate separate remuneration models and then to decompose the difference in mean log wages into parts attributable to differences in the means of the explanatory variables and to differences in the estimated coefficients:

$$\overline{\ln w^M} - \overline{\ln w^F} = (\bar{x}^M - \bar{x}^F)\hat{\beta}^M + \bar{x}^F(\hat{\beta}^M - \hat{\beta}^F) \qquad (1)$$

Here $\overline{\ln w}$ is the mean of the log of wages, $\bar{x}$ is the mean of the vector of explanatory variables, $\hat{\beta}$ is the vector of estimated coefficients from a Mincer log wage regression (Mincer, 1970; 1974) and the superscripts indicate gender (Oaxaca, 1973; Blinder, 1973).

The first part of the decomposition, due to differences in the means of the explanatory variables, is the explained wage difference and the second part, due to differences in coefficients, is the unexplained difference. The percentage of the overall difference due to the unexplained difference is used as a summary measure of discrimination since men and women with similar observed characteristics are rewarded differently for their characteristics.

The formulation is based on the implicit assumption that in the absence of discrimination women would be rewarded for their labour market characteristics at the same rate as men.[1] Although there is no compelling theoretical reason for this assumption it is adopted in most studies of gender discrimination. We follow the convention in most of our calculations. To illustrate the sensitivity of the estimate of discrimination to the assumed counterfactual, we also present some calculations based on the assumption that in the absence of discrimination men and women would be rewarded at the observed rate for women:

$$\overline{\ln w^M} - \overline{\ln w^F} = (\bar{x}^M - \bar{x}^F)\hat{\beta}^F + \bar{x}^M(\hat{\beta}^M - \hat{\beta}^F) \qquad (2)$$

Pay or log pay?

Some authors who have examined gender differences in physician pay, for example Kehrer (1976), Langwell (1982), estimate Mincer models with log wages as the dependent variable but then measure discrimination as the percentage difference in the mean level of wages. The counterfactual mean wage that female physicians would have earned if they were rewarded in the same way as males is estimated as $\exp(\bar{x}^F\hat{\beta}^M)$. However, as Ohsfeldt and Culler (1986) point out, $\exp(\bar{x}^F\hat{\beta}^M)$ is a biased estimate of the wage that a female with average female characteristics would have earned if rewarded as a male. They show that retaining the log wage specification but using alternative estimation methods to correct the bias leads to smaller estimates of gender discrimination.

In this paper, as in much of the recent work on gender discrimination, we prefer to measure discrimination as the difference in log wages, rather than the difference in wages. We thereby finesse the retransformation problem. Using the log difference permits much simpler estimation procedures. Moreover the log difference is also the only measure which satisfies compelling theoretical requirements for a measure of relative difference (Tornqvist, Vartia and Vartia, 1985)

Income or wages?

In most decomposition studies the measure of remuneration is the log hourly wage rate, with hourly pay calculated as total earnings divided by hours worked. If workers are on fixed hourly wages, so that pay is proportional to hours, then the procedure is sensible. But if pay is not proportional to hours, a remuneration function with the calculated wage as the dependent variable is misspecified. The coefficients in the estimated wage equation will reflect the market reward for characteristics such as experience, the effect of the characteristics on hours worked, and the effect of unobserved characteristics which influence hours worked.

---

[1] In general the difference in mean log incomes can be decomposed as

$$\overline{\log w}^M - \overline{\log w}^F = \left(\bar{x}^M - \bar{x}^F\right)\beta^* + \left[\bar{x}^M\left(\hat{\beta}^M - \beta^*\right) + \bar{x}^F\left(\beta^* - \hat{\beta}^F\right)\right]$$

where β^* is the vector of coefficients which would be obtained in a market in which there was no discrimination (Neumark, 1988). Neumark (1988) proposed that β^* be obtained as the coefficients from estimating the model $\log y_i = x_i\beta_i + \varepsilon_i$ on pooled data. Other suggestions can be taken as special cases of $\beta^* = \lambda\hat{\beta}^M + (1-\lambda)\hat{\beta}^F$. Oaxaca (1973) suggested that β^* could be either $\hat{\beta}^M$ or $\hat{\beta}^F$, i.e., λ =1 or λ = 0. Reimers (1983) suggested that λ = 0.5 and Cotton (1988) that $\lambda = n^1/\left(n^1 + n^2\right)$, where n^k is the number of individuals in group *k*.

If pay is not proportional to hours an earnings model which includes hours worked should be estimated. The decomposition can then be applied to the estimated earnings equations:

$$\begin{aligned}\overline{\ln y^M} - \overline{\ln y^F} &= (\bar{x}^M - \bar{x}^F)\hat{\beta}^{xM} + (\overline{\ln h^M} - \overline{\ln h^F})\hat{\beta}^{hM} \\ &\quad + \bar{x}^F(\hat{\beta}^{xM} - \hat{\beta}^{xF}) + \overline{\ln h^F}(\hat{\beta}^{hM} - \hat{\beta}^{hF}) \qquad (3)\end{aligned}$$

where y is earnings and h is hours.

Bashaw and Heywood (2001) show that because female physicians work fewer hours than male physicians and the marginal reward for hours declines with hours worked, the gender difference in mean physician log incomes (0.374) is much larger than the difference in mean log wages (0.146). After allowing for the difference in hours, the unexplained difference in log incomes is also much larger than the unexplained difference in log wages (0.177 versus 0.050). Following Bashaw and Heywood (2001) and studies of gender pay differences in the legal profession (McNabb and Wass, 2006; Wood, Corcoran and Courant, 1993) we focus on differences in income, rather than wages.

Endogeneity of hours.

Using income, rather than wages, as the dependent variable of interest raises an additional problem. Hours worked is an endogenous variable jointly determined with earnings and so the estimated coefficient on hours worked in the income regression may be biased. Since hours worked is an important determinant of earnings the bias in estimating the coefficients could have serious consequences for the decomposition of the earnings difference. Moreover, the coefficients on the exogenous variables in the income regression will also be biased if they are correlated with hours worked. Hence in this paper, unlike the previous literature (Bashaw and Heywood, 2001; McNabb and Wass, 2006; Wood, Corcoran and Courant, 1993), we use both OLS and two stage least squares to estimate the earnings equation, with marital status and number of children under 18 as instruments in the 2SLS model.

Dummy variables

The choice of reference category for dummy variables does not affect the size of the unexplained component and so does not affect the summary measure of discrimination. However, Oaxaca and Ransom (1999) show that the decomposition of the unexplained component in (1), (2), or (3) suffers from an identification problem if x includes one or more dummy variable. The attribution of the discrimination component to specific variables is then not invariant to the choice of reference category for the dummy variables. We adopt the solution proposed by Yun (2005) which is equivalent to calculating the decomposition with all possible regressions with different reference categories and averaging the results.

Regression constants.

When the x vector includes the unit constant, differences in the regression constant terms are counted as part of the unexplained difference. The constant in the regression is the average effect, conditional on included variables, of variables omitted from the regression. It depends on the unobserved coefficients on the omitted variables and that part of their unobserved mean which is not explained by the means of the included variables.[2] The gender difference in constant terms conflates differences in the unobserved coefficients on the omitted variables and the differences in their joint distributions with the included variables. Thus including the difference in regression constants will produce a biased estimate "unexplained" part of income differences, since part of the difference will be due to differences in distributions of characteristics, not differences in the rewards for characteristics. We provide a three way decomposition based on the regression model: explained, unexplained, difference in constants.

Insignificant gender differences in coefficients
The discrimination measure depends on the difference between coefficients. The usual practice is to estimate separate models and to report the percentage of the pay gap due to differences in coefficients from the two models. This procedure takes no account of the precision of the estimates of coefficients in the separate models. Thus large differences between imprecisely estimated coefficients can lead to a high estimate of discrimination.

There are two methods of allowing for the precision of the estimates of the coefficients. The first is to use the standard procedure and then to calculate the standard error for the "unexplained" component of the difference in mean pay. We do this using the *oaxaca* Stata module written by Jann (2005).

The second method, which is less usual in the literature is to estimate a model with restricted set of gender interactions, as in Wood, Corcoran, and Courant (1993). Although there are fewer coefficients which can differ, the estimated discrimination measure may increases or decrease since the excluded interaction terms are correlated with the remaining included variables. We compare the two approaches in our data set by comparing the estimated discrimination from a model estimated by dropping insignificant interaction terms with those from the model with a full set of gender interactions.

2.2 Matching methods

The regression based decomposition approach to analysing the difference in pay between male and female workers is vulnerable to two objections. First, it depends on the validity of the assumption that log income is linear in the explanatory variables. Second, the decomposition requires the counterfactual prediction of the income female doctors would have earned if they were rewarded for their characteristics in the same way as male doctors. It therefore assumes that the male doctor model is valid for the full range of the female explanatory variables, even though there may be no male doctors with explanatory variables over some parts of the range.

These criticisms are precisely those which have been levelled at regression based treatment effect models and so, by analogy, an alternative decomposition method

[2] Suppose the true model for gender k is $y_i^k = \beta_0^k + \beta^k x_i^k + \alpha^k q_i^k + \varepsilon_i^k$ (where x is the observed explanatory, q the omitted variable, and ε the zero conditional mean error). Then the expected value of the OLS regression constant is $\mathrm{E}\hat{\beta}_0^k = \beta_0^k + \alpha^k b_0^k$, where $b_0^k = \overline{q}^k - \overline{x}^k \sigma_{qx}^k / \sigma_{xx}^k$ is the constant term in the linear regression of q on x and $\sigma_{qx}^k, \sigma_{xx}^k$ are the covariance and variances for gender k.

based on matching estimators (Rosenbaum and Rubin, 1983) has recently been suggested (Barsky et al, 2001; Nopo, 2004; Djurdjevic and Radyakin, 2005). By interpreting gender as a "treatment" matching methods can be applied immediately and the unexplained effect of gender on income is estimated as the effect of treatment on the treated.

The key assumption justifying the use of matching methods is that assignment to treatment is statistically independent of the outcome conditional on the covariates or that there is selection only on observables. Although gender is not chosen or assigned, the interpretation of the selection on observables assumption is that the distribution of unobserved factors which affect outcome (income) is the same for treated and untreated (women and men). This is a strong and untestable assumption (which also underlies the interpretation of the difference in the regression constant as a component of discrimination). By including a large set of covariates we can hope to make the assumption more plausible.

Let $\ln y = \ell(x,k)$, $k = M, F$ be the possibly non-linear function determining the conditional expected log income of GPs of gender k. The unconditional expected log income of gender k GPs is

$$\begin{aligned}\mathrm{E}\left(\ln y^k\right) &= \int_S \ell(x,k)g(x|k,S)\,dx\,\Pr\left[S,k\right] + \int_{\bar{S}} \ell(x,k)g(x|k,\bar{S})\,dx\,\Pr\left[\bar{S},k\right] \\ &= \int_S \ell(x,k)g(x|k,S)\,dx \\ &\qquad + \left[\int_{\bar{S}} \ell(x,k)g(x|k,\bar{S})\,dx - \int_S \ell(x,k)g(x|k,S)\,dx\right]\Pr\left[\bar{S},k\right] \quad (4)\end{aligned}$$

where S is the common support for the distributions of x for male and female GPs, $\bar{S}$ is its complement, $g\left(x|k,S\right)$, $g\left(x|k,\bar{S}\right)$ are the corresponding conditional density functions, and $\Pr\left[S,k\right]$, $\Pr\left[\bar{S},k\right] = 1 - \Pr\left[S,k\right]$ the probabilities that a GP of gender k has characteristics in or out of the common support.

The difference in unconditional expected log incomes is

$$\begin{aligned}&\mathrm{E}\left[\ln y^M\right] - \mathrm{E}\left[\ln y^F\right] \\ &\quad = \int_S \ell(x,M)g(x|M,S)\,dx - \int_S \ell(x,F)g(x|F,S)\,dx \\ &\qquad + \underbrace{\left[\int_{\bar{S}} \ell(x,M)g(x|M,\bar{S})\,dx - \int_S \ell(x,M)g(x|M,S)\,dx\right]\Pr\left[\bar{S},M\right]}_{d_M} \\ &\qquad + \underbrace{\left[\int_S \ell(x,F)g(x|F,S)\,dx - \int_{\bar{S}} \ell(x,F)g(x|F,\bar{S})\,dx\right]\Pr\left[\bar{S},F\right]}_{d_F} \quad (5)\end{aligned}$$

d_M is the difference in the expected income of male GPs in and out of the common support. Similarly d_F is the difference in expected income of female GPs in and out of the common support. Neither term, or any combination of their components, is informative about the difference in incomes of similar male and female GPs: if the supports of the male and female distributions are identical then $d_M + d_F = 0$.

But the first pair of terms in (5) which compares incomes over the common support does provide useful information on the pay of similar male and female GPs. We can decompose it as

$$\int_S \ell(x,M)g(x|M,S)\,dx - \int_S \ell(x,F)g(x|F,S)\,dx$$

$$= \underbrace{\int_S \left[\ell(x,M) - \ell(x,F) \right] g(x|F)\, dx}_{d_o} + \underbrace{\int_S \ell(x,M) \left[g(x|M) - g(x|F) \right] dx}_{d_x} \quad (6)$$

which is a non-linear version of the Blinder-Oaxaca decomposition. d_o is the unobserved part of the difference in log incomes over the common support which arising from differences in the rewards for characteristics. d_x is the part due to differences in the distribution of male and female characteristics over the common support.

We calculate d_o without imposing any functional form on $\ell(x,M)$. We select female GPs one at a time and match them to a male GP or group of male GPs with similar observable characteristics and compute $\ell(x,M) - \ell(x,F)$ as the difference between the log income (or mean log incomes) of the comparator male GP (or GPs) and the female GP. The unweighted average of the differences is d_o: the unexplained part of the overall income difference which is not due to observable characteristics (since we have matched each female GP on similar male GPs.) It is the effect of the treatment (being female rather than male) on the treated (female GPs).

We use propensity score methods to find matching male GPs for female GPs. We estimate a probit model for being female using all observable practice and GP characteristics. We then match female GPs to male GPs with a similar estimated probability or propensity score. We use four variants of propensity score matching which we implement by adapting the *pscore* Stata module written by Becker and Ichino (2002).

Stratification matching: The stratification method is based on an algorithm which divides the individuals into a pre-specified number of strata depending on their propensity score. If the characteristics of the individuals in a given stratum are found to be significantly different the stratum is subdivided into two new strata, based on the propensity score. The algorithm continues until there are no significant differences between the characteristics of the individuals in all the strata. Female and male GPs may be in or out of the common support.

Nearest neighbour matching: The nearest neighbour method matches all female GPs with the male GP with the closest propensity score. All female GPs - except those whose propensity scores lie outside the intersection of the male and female propensity scores - are thus in the common support by construction, while some male GPs may not be. The matching is done with replacement which implies that the male GPs may be matched with more than one female GP.

Radius matching: The radius method matches female GPs with male GPs whose propensity scores are within a pre-defined range. Both male and female GPs may be in or out of the common support.

Kernel matching: The kernel method matches each female GP with all male GPs but attaches weights to the matches depending on the difference in the propensity scores.

With all the matching variants we treat the observations that do not belong to the intersection of the ranges of the male and female propensity scores as being out of the common support.

We compare the results from the various matching methods to those obtained using the standard regression-based Oaxaca-Blinder decomposition.

3 Family doctor contracts

The NHS is financed almost entirely from general taxation and patients face almost no charges for NHS health care. Patients register with general practices, which act as gatekeepers for elective hospital care. Their GPs are independent contractors, apart from a small minority who are directly employed by their local primary care organisation. They are organised in partnerships with a mean size of 3.7 GPs and around 6,200 patients. In 2004 40% of GPs were women.

Around two thirds of GPs are in practices with a nationally negotiated General Medical Services (GMS) contract under which they are paid by a mixture of capitation, lump sum allowances, items of service and target incentives. The capitation payments vary with the age of patients and with the deprivation level of the area in which they live. GPs have to meet all their practice expenses from their gross income, except for some specific reimbursements for the costs of practice nurses and computing systems. Additionally, where there is no local pharmacy, GPs are permitted to dispense the medicines they prescribe. Dispensing practices can make a profit from dispensing since they receive a dispensing fee per item and are reimbursed for the drugs they buy at a rate that often exceeds the price they paid.

Around one third of practices have opted to be paid under a Primary Medical Services (PMS) contract. These contracts are negotiated between the practice and their local primary care organisation. Under the PMS contract, the practice receives a lump sum in exchange for agreeing to provide the services they would have provided under the GMS contract, plus additional services for particular patient groups. The amount received is typically the amount the practice would have received under GMS, plus an addition intended to cover the cost of the extra services. As under GMS, the practice has to meet its expenses from its gross income.

4 Data

The data are from the 2004 National Primary Care Research and Development Centre's General Practitioner Worklife survey (Whalley et al, 2005). The questionnaire was posted in February 2004 to 4208 salaried and principal GPs in England. We use responses from 598 female and 1178 male GPs. The respondents were asked to report their banded[3] earnings, defined as total annual income from their practice before taxes but after deducting expenses from their practice; usual hours worked per week; and personal and practice characteristics.

[3] The earnings bands are: less than £25,000, £25,000-£49,999, £50,000-£69,999, £70,000-£84,999, £85,000-£99,999, £120,000-£149,999 and £150,000 or more.

We were also able to add further information regarding the GPs' practices from the Attribution Data Set which has detailed information on variables which determine the practice’s capitation income and from the General Medical Services data set maintained by the Department of Health.

We divide the variables potentially affecting GP remuneration into:

Personal characteristics. We include a measure of experience (decades since qualification) and its square to allow for the usual positive but declining effect of experience on wages and earnings. We include ethnicity (white/non-white) to allow for possible discrimination based on race. We also know whether the GP qualified in the UK or abroad. This will be correlated, though not perfectly with ethnicity, and because we also include ethnicity the coefficient on non-UK qualified may reflect the impact of less relevant initial experience on earnings. We include an indicator for salaried status and expect that it will have a negative effect on earnings.

We include an indicator for whether the GP is the senior partner in the practice. Many general practice partnerships own their premises and so some of the payment a GP receives may be a return on capital invested in the practice. The senior partner is likely to contribute the largest proportion of the capital invested. We expect that senior partners will have higher incomes than otherwise similar GPs.[4]

We use a set of family characteristics as instruments for hours worked in the 2SLS models of earnings. We know the GP’s marital status and whether their partner works. We collapse this information into a single dummy which takes the value 1 if the GP has a partner who does not work and 0 otherwise (no partner or partner works). We expect that GPs with a non working partner will work longer hours to boost family income and because a non-working partner will be able to provide child care. We also include dummy variables for numbers of children under 18.

Practice characteristics. Under the GMS contract the per patient capitation income of the practice depends on three adjustments. The age/sex adjustment is a weighted sum of the age and sex mix of the practice population, with the weights intended to reflect national age and sex specific consultation rates. Broadly, the practice receives more gross income the older its population. The nursing home adjustment is based on the proportion of the local population in nursing homes, with higher capitation paid the higher the proportion. The additional needs adjustment is based on the morbidity of the population as measured by the standardised limiting long term illness ratio and the standardised mortality ratio. We have information on these three adjustments for each practice from the Attribution Data Set. The nursing home adjustment and the additional need adjustment are attributed to practices using the proportion of their patients living in different areas

We include the number of practice patients per whole time equivalent GP and expect GPs in practices with larger lists per GP to have higher earnings. We also include the

[4] We also had a measure of the length of time the GP had been in the practice. It is highly correlated with experience and senior partner status and including it in the model made no difference to the results.

total practice population. The total list is highly correlated with number of GPs and the incentive literature (Gaynor and Pauly, 1990) suggests that practices with more GPs will have lower earnings per GP. On the other hand, there may be economies of scale in delivering services, so that earnings per GP may increase with total list size.

Practices can earn additional income from patients by providing them with various fee per item services, such a flu vaccinations, and by meeting targets for vaccination of children and cervical screening. It may be harder to generate this type of income from less educated or poorer populations. We include a measure of the education of the local population and the Low Income Scheme Index score. The LISI score is the cost weighted proportion of practice prescriptions which are dispensed without charge to the patient on grounds of low income (Lloyd, Harris and Clucas, 1995).

We include indicators of whether the practice has a PMS contract and whether it is permitted to dispense, as well as prescribe, pharmaceuticals to its patients. Given the relative information and incentives of GPs and PCT managers we expect PMS practices to have negotiated a contract with their PCT which generates higher total practice profits than they would earn under the nationally negotiated GMS contract. Dispensing practices are reimbursed for their purchases of medicines at fixed prices which generally exceed the actual purchase price and so GPs in dispensing practices should have higher incomes.

We also include an indicator of whether the practice elected to hold a budget for elective inpatient expenditure under the fundholding scheme which ran from 1991/2 to 1998/9. GPs in fundholding practices have been shown to be more entrepreneurial and income orientated than non-fundholders (Whynes Ennew and Feigham, 1999). Thus we expect that GPs in such practices will have higher incomes..

Regional dummies. We use 8 Government Office Region dummies to capture, inter alia, geographical differences in the cost of practice non-GP inputs, such as practice nurses.

5 Measuring gender discrimination: results

5.1 Descriptive statistics

The sample has broadly similar characteristics to the GP population (Department of Health, 2004) in terms of gender (34% female against 38% in the GP population), PMS status (33% against 37%), and working in a dispensing practice (18% against 16%).

Table 1 has descriptive statistics for the estimation sample. The mean annual gross earnings for male respondents is £81,288 while the mean earnings among female respondents is £57,312 or about 71% of mean male earnings.[5] Male GPs work 48.4 hours per week and female GPs 36.7 hours per week on average, or about 76% of

[5] Based on the midpoint of the reported income band. The 0.7% of the respondents earning £150,000 or more are classified as earning £175,000.

male hours. Female GPs have a hourly wage which is 92% of the male wage, where the hourly wage is calculated as income/(weekly hours*47).

Male GPs have slightly more experience than female GPs (22.9 vs. 20.1 years) and are more likely to be non-white, qualified outside the UK and on a salaried contract. They are also more likely to be married with a non-working partner but there is no substantial difference in the likelihood of not having children.

Male GPs are in practices have a larger total list and a larger list per GP. While male GPs tend to work in practices with a higher share of patients living in nursing homes, male and female GPs work in similar practices in terms of the patients' age/sex mix and degree of morbidity as measured by the additional needs adjustment. Male GPs tend to work in practices with somewhat smaller LISI scores but are more likely to work in a practice located in a ward with a high share of people with no qualifications. They are also more likely to work in dispensing, PMS, and ex-fundholder practices.

5.2 Regression models

We pool the male and female samples and first report estimates of OLS models with a full set of gender interactions

$$z_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + F_i\left(\alpha_0 + x_{1i}\alpha_1 + x_{2i}\alpha_2\right) + \varepsilon_i \qquad (7)$$

and heteroscedasticity robust standard errors. Here z_i is either log wage or log income (Table 2, columns 2 and 3).[6]

The results for the coefficients which are common across the log wage and the log earnings models are qualitatively similar in terms of signs and significance. GPs in practices with larger lists have higher wages and earn more: the effects of economies of scale seem to more than offset the attenuation of incentives. The number of patients per GP has a positive but insignificant effect except for female wages where it has a negative and insignificant effect. A higher LISI score, indicating a more deprived population, is associated with lower wages and earnings. The contractual age/sex adjustment has a negative effect. Since the contract gives practices a higher average capitation payment the higher its age/sex adjustment, these results suggest that the higher revenue is more than offset by higher costs associated with more demanding populations. The effects of the nursing home and additional needs adjustments are positive and insignificant except for female wages where the nursing home effect is negative and insignificant.

Being in a dispensing practice or a PMS practice has large positive and significant effects on wages and earnings. Working in an ex-fundholding practice has a positive effect for both male and female GPs, but the effect is only significant for male GPs. Salaried GPs have lower wages/earnings than other GPs but the effect is not significant. Senior partners earn more than other GPs. Since we also control for experience and hours, this suggests that the higher rewards for senior partners reflect larger unobserved investment by them.

---

6 Both the wage and income variables are based on the midpoints of the reported income bands. We also estimated interval regression models (e.g. Jones, 2000) which gave very similar results and are therefore not reported.

In none of the models is experience or its square significant. It may be that years since qualification is a poor measure of relevant experience, especially for GPs with children, or GPs from overseas. We attempted to allow for this possibility by interacting the experience variable with a dummy for having children, but this did not lead to a qualitative change in the results. Being non-white has a very small negative and insignificant effect on male GP wages/earnings but a larger positive and significant effect on female GP earnings.

The coefficients for log hours in the log earnings regressions in Table 2 are significantly less than one for both male and female GPs, suggesting that earnings are not proportional to hours worked. A 1% increase in hours worked leads to a 0.25% increase in earnings for male GPs and a 0.56% increase for female GPs.

Table 3 shows the relationship between predicted income, wages and hours worked. Predicted income is derived by estimating a levels model of income separately for male and female GPs. Figure 1 plots estimated income against hours and Figure 2 the implied hourly wage against hours.

Studies of other professional groups also find that the marginal returns to hours worked declines with hours worked. Gaynor and Pauly (1990) report an elasticity of physician output (office visits) with respect to hours of 0.53. Conrad et al (2002) find an elasticity of total charges with respect to hours of 0.46. Bashaw and Heywood (2001) for US physicians, find the elasticity is 0.25 for male physicians and 0.44 for female physicians. In McNabb and Wass (2006) the elasticity of British solicitors' (lawyers') income with respect to hours is 0.29 for men and 0.58 for women.

Because the wage declines with hours worked and female GPs work fewer hours, the gender difference in average wages (or in average log wages) will be smaller than the difference in average incomes (or average log incomes). More importantly it implies that log wage models are misspecified because of their implicit assumption that earnings are proportional to hours. Thus differences in pay should be investigated via models of income, not wages.

Table 2 (columns 2 and 3) reports results from OLS models of log income. To test for the endogeneity of hours we also estimated a 2SLS earnings model, using dummy variables for the number of children in the respondent's household and whether the respondent has a working partner as instruments for hours worked and hours worked interacted with the dummy for being female. The instrumental variables were jointly significant ($F=9.78$, $p<0.001$) in the first-stage (hours) regressions, indicating that the instruments satisfy one of the requirements for validity. The models also passed the robust Sargan test for overidentifying restrictions (Baum, Schaffer and Stillman, 2003), suggesting that the instruments are uncorrelated with the error term ($\chi^2=9.78$, $p=0.96$).

A regression-based Hausman test (Wooldridge 2002, pp. 121) could not reject the null hypothesis of hours worked being exogenous in the earnings regression ($F=0.30$, $p=0.74$), however, and we therefore report the OLS results only.[7] The results from the

[7] Conrad et al (2002) test for endogeneity of hours worked and also find that hours are not endogenous and that OLS and 2SLS estimates of the effect of hours on output are very similar.

2SLS regressions are very similar to those from the OLS models and available from the authors upon request.

We also estimated an income model in which we directly controlled for the number of children in the respondent's household and whether the respondent has a working partner. The effects of these characteristics were jointly insignificant for both male (F=0.36, p=0.84) and female GPs (F=0.33, p=0.86). This is in line with the finding by Sasser (2005) that once hours worked is controlled for the effect of having children on earnings becomes insignificant, suggesting that while having children influences earnings through a reduction in hours worked, it does not have a direct influence on earnings. This corresponds to the assumptions underlying the 2SLS model.

Most of the gender interaction terms in the log income model are individually insignificant. We therefore estimated a more parsimonious OLS model retaining only interaction terms on log hours, experience, experience squared and ethnicity:

$$\ln y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + F_i\left(\alpha_0 + x_{1i}\alpha_1\right) + \varepsilon_i \tag{8}$$

The results are reported in Table 2, column 4. The dropped interaction terms were jointly insignificant at the 5% level (F=1.18, p=0.26). The coefficients on the variables included in the reduced model are very similar to those in the model with the full set of interactions.

5.3 Regression based measures of discrimination

Table 4 reports the results of the decompositions of the differences in mean log wages and income from the models in Table 2. The difference in mean log wages is 0.084 log-points and highly significant (t=4.79, p<0.001), as is the much larger relative difference in mean log income of 0.384 log-points (t=25.21, p<0.001).

The unexplained component of these differences is larger when the counterfactual is what female GPs would have earned if they had been rewarded as if they were male GPs $\left(\bar{x}^F \hat{\beta}^M\right)$. For example, the unexplained log percentage difference in the log income model with a full set of interactions is 28% using the counterfactual $\left(\bar{x}^F \hat{\beta}^M\right)$ and 19% when using the counterfactual $\left(\bar{x}^M \hat{\beta}^F\right)$. The differences between decompositions based on the log income and log wage models for a given counterfactual is much smaller.

Two other features of the decompositions are noteworthy. First, the overall unexplained part has both negative and positive components which are larger in absolute value than the overall difference in log wage or income. In the log income case female GPs get smaller rewards for personal characteristics (experience, experience squared) than male GPs but higher rewards for hours and practice characteristics. Second, the largest component of the unexplained part of the difference in mean log income is the constant term. Since, as we suggested in section 2, the difference in the constant terms reflects both the difference in rewards attached to unobserved variables and differences in the means of these variable, the interpretation of the unobserved component when it includes the difference in constants as discrimination is questionable.

The log income decomposition in Table 4 (column 4) is based on the model (Table 2, column 2) with the full set of interactions. Decompositions based on the more parsimonious model (Table 2, column 3) gave very similar results. The explained and unexplained components 0.106 (t=6.09, p<0.001) and 0.278 (t=11.08, p<0.001) when the male coefficients are used as counterfactuals and 0.226 (t=11.11, p<0.001) and 0.158 (t=6.90, p<0.001) when the female coefficients are used as counterfactuals.

5.4 Matching based measures of discrimination

Tables 5 and 6 report the decompositions of differences in mean log income and mean income based on propensity score matching. The probit regression model used to generate the propensity scores is Appendix Table 1. As the summary statistics in Table 1 suggest, hours and family circumstances have a powerful effect on the prediction of gender. Figure 3 plots the kernel densities for the distributions the predicted probability of being female for female and male GPs. It can be seen from the figure that there is a considerable degree of overlap between the two distributions. Only 9 GPs, who are all male, have propensity scores outside the overlapping range [.0010, .9654].

Tables 5 and 6 show that the different matching methods produce similar estimates of the average treatment effect (d_o), which is a common finding in the matching literature (Smith and Todd, 2005).[8] The treatment effect is the effect of being female on earnings net of any differences in endowments or workplace attributes. The log percentage difference in mean log income ranges from 21% to 27%. The treatment effect on mean income ranges from £13,800 to £16,700.

The treatment effect is the non-parametric analogue of the unexplained component in the Oaxaca decomposition, which in the log income model is a log percentage of 28%. There is therefore agreement between the parametric and non-parametric estimates of the extent of the difference in log incomes which is not explained by observable characteristics of male and female GPs.

The matching models are less consistent in the estimates of the remaining components of the wage gap; the endowment effect and the components of the decomposition attributed to the differences between male and female GPs in and out of the common support. This is unsurprising since the assumptions underlying the different approaches differ markedly in this respect. The stratification and kernel methods find a match for all the individuals in the sample - with the exception of the 9 male GPs who have propensity scores outside of the overlapping range of male and female propensity scores which we consider being out of the common support in all the approaches - but differ in the way they attribute weights to the matches. The nearest neighbour method finds a match for all female GPs, but not all male GPs are matched with a female. The radius method only matches female GPs with male GPs whose propensity scores are within the pre-specified range (set to 0.01), with the result that no match could be found for some male and female GPs in the sample.[9] As no

[8] The bootstrap t statistics are from 500 replications. Abadie and Imbens (2006) warn that the bootstrapped variance is not reliable in the case of nearest neighbour matching.

[9] The radius method is very sensitive to the specified radius.

matching method is *a priori* superior we do not attempt to draw any firm conclusions as to the proportion of the wage gap attributed to differences in endowments between male and female GPs in the common support and the proportion attributed to differences between GPs in and out of the common support.

6 Testing for within practice discrimination

6.1 A direct test for within practice gender discrimination

The pay of GPs is determined by total practice income and by the within practice income sharing rules. We have shown that there are statistically and economically significant differences in pay between male and female GPs. Such pay differences may be due to paying otherwise identical individuals different amounts or they may be due to paying GPs according to their income generating activity and there being unobserved differences in these activities.

GPs produce a wide range of services which have different effects on the well being of their patients and on the income of the practice. There is some evidence that male and female doctors produce different output mixes and have different preferences. Langwell (1982) found that female physicians saw fewer patients per hour. Female GPs have longer consultations (Wilson, 1991). Rizzo and Zeckhauser (2007) report differences in male and female attitudes to income generation. More generally, experiments suggest different attitudes to competition and cooperation in teams (Niederle and Vesterlund, 2005; Ivanova-Stenzel and Kubler, 2005).

Within practice discrimination can take three forms. Female GPs may be assigned a mix of activities which generate less income. The practice income sharing formula may give lower rewards to activities which women prefer or in which they have a comparative advantage. Or women may get a smaller reward for any given mix of activities.

None of these types of within practice discrimination can be present in practices where all the GPs have the same gender. Hence differences in the incomes of female and male GPs working in single gender practices, conditional on the exogenous factors affecting total practice income, must be due to differences in preferences or productivity in income generation. Subtracting this difference from the difference in income between female and male GPs working in mixed gender practices provides an estimate of the discrimination: the difference in incomes not due to gender differences in productivity or preferences.

We estimate the following model for GP income:

$$\ln y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + F_i\left(\alpha_0 + x_{1i}\alpha_1\right) + U_i\left[\gamma_0 + x_{1i}\gamma_1 + F_i\left(\delta_0 + x_{1i}\delta_1\right)\right] + \varepsilon_i \qquad (9)$$

where $F_i = 1$ if the GP is female, $U_i = 1$ if the practice GPs are all of the same gender, x_1 is vector of variables whose effects vary with the gender of the GP and whether all GPs in the practice are of the same gender, and x_2 is a vector of covariates whose effects are the same for all GPs. Using the results in section 5.2, we assume that x_1, the subset of variables whose effects vary by gender, consists only log hours, experience, experience squared and ethnicity.

We use the results from estimating (9) for three direct tests of within practice gender discrimination.

(i) The conditional mean difference in log income between females in female only and mixed practices is

$$y_{ff} - y_{fm} \equiv \gamma_0 + \delta_0 + \bar{x}_1^F(\gamma_1 + \delta_1) \qquad (10)$$

Whatever the relative productivity or taste differences of male and female GPs pro male discrimination implies that, conditional on their other characteristics, females will get higher income in female only practices than in mixed practices. (And conversely if there is pro-female discrimination.) Thus pro-male discrimination implies that $y_{ff} - y_{fm} > 0$.

(ii) The conditional mean difference in log income between males in mixed and male only practices is

$$y_{mm} - y_{mf} \equiv \gamma_0 + \bar{x}_1^M \gamma_1 \qquad (11)$$

Whatever the relative productivity or taste differences of male and female GPs pro male discrimination implies that, conditional on their other characteristics, male GPs will get lower income in male only practices than in mixed practices: there are no females to exploit in male only practices. (Conversely if there is pro-female discrimination.) Thus pro-male discrimination implies that $y_{mm} - y_{mf} < 0$.

(iii) The conditional mean difference in log income between a female in an all female practice and a male in an all male practice is

$$[\beta_0 + \bar{x}_1^F\beta_1 + \bar{x}_2^F\beta_2 + \gamma_0 + \bar{x}_1^F\gamma_1 + \alpha_0 + \bar{x}_1^F\gamma_1 + \delta_0 + \bar{x}_1^F\delta_1]$$
$$- [\beta_0 + \bar{x}_1^M\beta_1 + \bar{x}_2^M\beta_2 + \gamma_0 + \bar{x}_1^M\gamma_1]$$

$$= \left[\alpha_0 + \delta_0 + \bar{x}_1^F(\alpha_1 + \delta_1)\right] - \left[\Delta\bar{x}_1\beta_1 + \Delta\bar{x}_2\beta_2 + \Delta\bar{x}_1\gamma_1\right] \qquad (12)$$

The second term in (12) is due entirely to differences in average male and female characteristics. The first term is due to differences in coefficients showing the effect of being female (α_0, α_1) rather than male in a mixed sex practice and of being female rather than male in a unisex practice (δ_0, δ_1). The first term shows the effect of changing sex and practice gender mix on someone who has average female characteristics. The difference in log income for a GP who is female rather than male (but with all other characteristics equal to those of the average female GP) in a single sex practice is

$$y_{ff} - y_{mm} \equiv \left[\alpha_0 + \delta_0 + \bar{x}_1^F(\alpha_1 + \delta_1)\right] \qquad (13)$$

Thus $y_{ff} - y_{mm}$ measures the difference in log income due to productivity or taste differences rather than to discrimination.

The conditional mean difference between a female in a mixed gender practice and a male in a mixed gender practice, each with otherwise average female characteristics is

$$y_{fm} - y_{mf} \equiv \left[\alpha_0 + \bar{x}_1^F\alpha_1\right] \qquad (14)$$

This difference may due to taste or productivity differences as well as pro-male discrimination. Using (13) the extent of pro-male discrimination is measured by

$$(y_{fm} - y_{mf}) - (y_{ff} - y_{mm}) = -(\delta_0 + \bar{x}_1^F\delta_1) \qquad (15)$$

The income difference between all male and all female practices may reflect selection of men into practices with unobservable practice factors making the practice more profitable. We cannot rule out discrimination in the selection of female GPs into practices, so that some of the difference in incomes in all male and all female practices in (13) may be due to discrimination in selection rather than to differences in within practice productivity or tastes. However, we have a rich set of practice variables, some of them from the national capitation formula used to determine practice income, and so hope that there are no unobservable factors affecting practice income and correlated with GP gender.

6.2 Testing for within practice discrimination: results

We estimate two versions of (9). In the first we set γ_1 and δ_1 to zero so that the effect of being in unisex practice on female and male GP income is assumed not to vary with any other characteristics. In the second version we permit $\gamma_1 \neq 0$ and $\delta_1 \neq 0$ so that the effect of the gender mix in the practice on female and male GP income may also depend on those characteristics which have been previously shown to have differential effects on female GP income.

Since small practices are more likely to have all GPs of the same gender we estimate the model with a quartic functions of total list size and of list size per GP. This will reduce the risk that the estimated effect of the unisex variable is contaminated by any misspecification of the effects of size on income.

Table 7 has the regression results from the two variants of (9) with the flexible size specification. The coefficients on the covariates other than female, unisex and their interactions with log hours, experience, experience squared and ethnicity are very similar. The coefficients are also similar to those in Table 2. Table 8 reports the direct tests for productivity and taste differences and discrimination based on the two regressions in Table 7.

(i) Discrimination against female GPs would imply that they would earn more in all female practices than in mixed practices ($y_{ff} - y_{fm} > 0$). The differences in conditional mean log incomes in the two models are positive and reasonably large (0.08, 0.05) compared with the overall unconditional income difference of 0.38, but not close to conventional levels of statistical significance ($p = 0.27, 0.56$).

(ii) Both models show that GPs in all male practices earn more than those in mixed practices ($y_{mm} - y_{mf} > 0$), suggesting, if anything, that there is pro-female discrimination. However, the differences in conditional mean log incomes (0.03, 0.03) are not significant ($p = 0.29, 0.25$).

(iii) In both models the difference between the conditional mean log incomes of female and male GPs in mixed practices ($y_{fm} - y_{mf}$) is negative, large (–0.28, –0.27), and highly significant ($p < 0.001$). But the conditional mean differences between the log incomes of female GPs in all female practices and male GPs in all male practices ($y_{ff} - y_{mm}$) are also negative, large (–0.23, –0.31) and statistically significant ($p = 0.001$). After allowing for these differences due to productivity or taste differences the conditional mean difference in log incomes between female and male GPs in mixed practices $(y_{fm} - y_{mf}) - (y_{ff} - y_{mm})$ is still negative in the model with a more limited set of interactions but it is greatly reduced (–0.05) and statistically insignificant (p =0.51). For the model with a fuller set of interactions, female GPs have larger log incomes than male GPs after allowing for taste and productivity differences: $(y_{fm} - y_{mf}) - (y_{ff} - y_{mm}) = 0.05$ (p =0.63).

The results from the three direct tests suggest that there is no within practice discrimination against female GPs.

7 Conclusions

Conditional on observed factors, female GPs have lower incomes, lower average wages (income/hours) but higher marginal rewards for hours. The elasticities of income with respect to hours are significantly less than 1 and much larger for women than for men. Thus models of GP pay based on wages calculated as income/hours are misspecified. Since female GPs work shorter hours (37 per week against 48), the relative log gender difference in wages (8%) is markedly less than the relative log gender difference in incomes (38%).

The unexplained difference in log incomes is 28% and is insensitive to the specification of the log income regression. Estimates of discrimination based on comparisons of log income for GPs with similar propensity scores yield slightly smaller measures (21% to 27%).

Using our direct tests for within practice income discrimination based on the comparison of GPs in practices with differing gender mixes we cannot reject the null hypothesis of no gender discrimination. The tests suggest that unexplained differences in pay are due to differences in unobservable tastes and productivity concerning GP income generating activity.

References

Abadie, A.and Imbens, G. 2006. On the failure of the bootstrap for matching estimators. *National Bureau of Economic Research, Technical Working Paper* 325.

Ash, A.S., Carr, P.L., Goldstein, R. and Friedman, R.H. 2004. Compensation and advancement of women in academic medicine: is there equity? *Annals of Internal Medicine*, 141, 205-212.

Baker, L.C. 1996. Differences in earnings between male and female physicians. *New England Journal of Medicin*e, 334, 960-964.

Barsky, R., Bound, J., Charles, K. and Lupton, J.P. 2002. Accounting for the black-white wealth gap: a non-parametric approach. *Journal of the American Statistical Association*, 97, 663-673.

Bashaw, D.J. and Heywood, J.S. 2001. The gender earnings gap for US physicians: has equality been achieved? *Labour*, 15, 371-391.

Baum, C.F., Schaffer, M.E. and Stillman, S. 2003. Instrumental variables and GMM: Estimation and testing. *The Stata Journal* 3, 1-31.

Becker, S.O. and Ichino, A. 2002. Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2, 358-377.

Blackaby, D., Booth, A.L. and Frank, J. 2005. Outside offers and the gender pay gap: empirical evidence from the UK academic labour market. *Economic Journal*, 115, F81-107.

Blinder A. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources* 8, 436-455.

Conrad, D.A., Sales, A., Liang, S.Y., Chaudhuri, A., Maynard, C., Lieper, L., Weinstein, L., Gans, D. and Piland, N. 2002. The impact of financial incentives on physician productivity in medical groups. *Health Services Research*, 37, 885-906.

Cotton J. 1988. On the decomposition of wage differentials. *Review of Economics and Statistics* 70, 236-243.

Department of Health. *Statistics for General Medical Practitioners in England 1994-2004.* 2005.

Djurdjevic, D. and Radyakin, S. 2005. Decomposition of the gender gap using matching: an application for Switzerland. *Darmstadt Discussion Papers in Economics* 155.

Dolton, P., O'Neil, D. and Sweetman, O. 1996. Gender differences in the changing labor market: the role of legislation and inequality in changing the wage gap for qualified workers in the United Kingdom. *Journal of Human Resources*, 31, 549-565.

Gaynor, M and Pauly, M.V. 1990. Compensation and productive efficiency in partnerships: evidence from medical groups practice. *Journal of Political Economy*, 98, 544-573.

Gupta, N., Diallo, K., Zurn, P., and Poz, M.R.D. 2003. Assessing human resources for health: what can be learned from labour force surveys?. *Human Resources for Health* 1, http://www.human-resources-health.com/content/1/1/5

Ivanova-Stenzel, R. and Kubler, D. 2005. Courtesy and idleness: gender differences in team work and team competition. *Humbold University Discussion Paper SFB* 2005-049.

Jann, B. 2005. Standard errors for the Blinder-Oaxaca decomposition. Presented at the 3rd German Stata Users Group Meeting, Berlin.

Jones, A. 2000. Health Econometrics. In: Culyer, A.J., Newhouse, J.P. (Eds.), *Handbook of Health Economics*. Elsevier, Amsterdam.

Kehrer, B.H. 1976. Factors affecting the incomes of men and women physicians: an exploratory analysis. *Journal of Human Resources*. 11, 526-545.

Langwell, K.M. 1982. Factors affecting the incomes of men and women physicians: further explorations. *Journal of Human Resources*, 17, 261-275.

Lloyd, D.C.E.F., Harris, C.M., and Clucas, D.W. 1995. Low income scheme index: a new deprivation scale based on prescribing in general practice. *British Medical Journal*, 310, 165-169.

McNabb, R. and Wass, V. 2006. Male-female earnings differentials among lawyers in Britain: a legacy of the law or current practice. *Labour Economics* 13, 219-235.

Mincer J. 1970. The distribution of labour incomes: A survey with special reference to the human capital approach. *Journal of Economic Literature* 8, 1-26.

Mincer J. 1974. *Schooling, Experience and Earnings*. National Bureau of Economic Research. Columbia University Press, New York.

Niederle, M. and Vesterlund, L. 2005. Do women shy away from competition? Do men compete too much?. *NBER Working Paper* 11474.

Neumark D. 1988. Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23, 279-295.

Noonan, M., Corcoran, M. and Courant, P. 2005. Pay differences among the highly trained: cohort differences in the sex gap in lawyers' earnings. *Social Forces*, 84, 853-872

Nopo, H. 2004. Matching as a tool to decompose wage gaps. *IZA Discussion Paper* 981.

Oaxaca R. 1973. Male-female wage differentials in urban labour markets. *International Economic Review* 14, 693-709.

Oaxaca, R.L., and Ransom M.R. 1999. Identification in detailed wage decompositions. *Review of Economics and Statistics* 81, 154–57.

Ohsfeldt, R.L. and Culler, S.D. 1986. Differences in income between male and female physicians. *Journal of Health Economics*, 5, 335-346.

Reimers C. 1983. Labor Market discrimination against hispanic and black men. *Review of Economics and Statistics* 65, 570-579.

Rizzo, J.A. and Zeckhauser, R.J. 2007. Pushing incomes to reference points: why do male doctors earn more? *Journal of Economic Behaviour and Organization*, 63, 514-536.

Robinson, D. 1998. Differences in occupational earnings by sex. *International Labour Review*, 137, 3-31.

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

Sasser, A.C. 2005. Gender differences in physician pay. *Journal of Human Resources*, 40, 477-504.

Smith, J.A. and Todd, P.E. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics* 125, 305-353.

Tornqvist, L., Vartia, P. and Vartia, Y., 1985. How should relative changes be measured? *The American Statistician*, 39, 43–46.

Whalley, D., Bojke, C., Gravelle, H. and Sibbald, B. 2004. *2004 National Survey of General Practitioner Job Satisfaction. Final Report.* December www.npcrdc.man.ac.uk/Publications/GP%20Satisfaction%20REPORT.pdf?CFID=792808&CFTOKEN=37289229

Whynes, D., Ennew, C. and Feigham, T. 1999. Entrepreneurial attitudes of primary health care physicians in the United Kingdom. *Journal of Economic Behavior and Organization*, 38, 331-347.

Wilson, A. 1991. Consultation length in general practice: a review. *British Journal of General Practice*. 41, 119-122.

Wood, R.G., Corcoran, M.E. and Courant, P.N. 1993. Pay differences among the highly paid: the male-female earnings gap in lawyers' salaries. *Journal of Labour Economics*, 11, 417-441.

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge Massachusetts, USA

Yun, M-S. 2005. A simple solution to the identification problem in detailed wage compositions. *Economic Inquiry*, 43, 766-772.

Table 1. Descriptive statistics

Variable	Male GPs Mean	Male GPs SD	Female GPs Mean	Female GPs SD	
Annual earnings in £`000s	81.288	22.669	57.312	22.244	**
Weekly hours worked	48.407	11.173	36.664	11.938	**
Decades of experience	2.293	0.697	2.010	0.707	**
Non-white	0.158		0.112		**
Non-UK qualified	0.121		0.089		*
Salaried GP	0.028		0.020		
Senior GP	0.374		0.152		**
Married – partner does not work	0.229		0.067		**
No children	0.326		0.321		*
One child under 18	0.159		0.135		
Two children under 18	0.295		0.356		
Three or more children under 18	0.221		0.187		
Low Income Scheme Index (LISI) score	9.295	5.886	9.453	6.179	
Proportion in ward with no qualifications	0.286	0.094	0.275	0.095	*
Practice list size in `000s	8.953	4.363	8.643	4.051	
Practice list size per WTE GP in `000s	2.024	0.471	1.953	0.378	**
Age/sex adjustment	22.42	0.914	22.281	0.879	**
Nursing home adjustment	24.87	24.035	15.203	15.623	**
Additional needs adjustment	96.56	10.722	96.558	10.808	
Dispensing practice	0.191		0.169		
PMS practice	0.342		0.321		
All practice GPs same gender	0.178		0.070		**
Ex-fundholder practice	0.545		0.508		
Government Office Region:					
GOR1 - North East	0.042		0.048		*
GOR2 - North West	0.133		0.132		
GOR3 - Yorkshire	0.092		0.114		
GOR4 - East Midlands	0.092		0.062		
GOR5 - West Midlands	0.094		0.080		
GOR6 - East	0.115		0.095		
GOR7 - London	0.120		0.159		
GOR8 - South East	0.171		0.192		
GOR9 - South West	0.142		0.117		
N	1178		598		

Significance levels for differences in means or proportions: *p<0.05, **p< 0.01; t tests except Chi square for numbers of children and GOR. The adjustments for age/sex, additional needs, and nursing homes are adjustments to the capitation payments. They reflect differences in the demographic mix of the practice patients, the socio-economic characteristics of the local population, and the proportion of the local population in nursing homes.

Table 2. Log wage and income regressions

Variable	Log wage model Coef.	Log wage model t-stat.	Log income model all interactions Coef.	Log income model all interactions t-stat.	Log Income model parsimonious Coef.	Log Income model parsimonious t-stat.
Log of weekly hours worked			0.246	4.64	0.251	4.70
Log of weekly hours worked x Female			0.321	3.69	0.329	3.88
Female	-0.345	-0.13	-2.033	-0.89	-1.240	-3.57
Experience in decades	0.000	0.01	0.111	1.32	0.103	1.25
Experience x Female	-0.179	-1.07	-0.296	-2.01	-0.290	-1.99
Experience squared	0.004	0.21	-0.027	-1.44	-0.026	-1.42
Experience squared x Female	0.040	1.02	0.077	2.23	0.079	2.30
Non-white	-0.042	-1.20	-0.011	-0.37	-0.020	-0.70
Non-white x Female	0.175	2.85	0.143	2.73	0.184	3.61
Non-UK qualified	-0.025	-0.69	0.006	0.19	-0.005	-0.19
Non-UK qualified x Female	0.018	0.25	-0.001	-0.02		
Salaried GP	-0.070	-1.44	-0.082	-1.55	-0.087	-2.02
Salaried GP x Female	-0.006	-0.05	-0.023	-0.26		
Senior GP	0.024	1.02	0.065	3.41	0.083	4.45
Senior GP x Female	0.050	0.86	0.064	1.22		
Log LISI score	-0.080	-2.57	-0.076	-3.27	-0.062	-3.00
Log LISI score x Female	0.039	0.74	0.037	0.81		
Log prop. in ward w/o qualifications	-0.014	-0.40	0.007	0.19	-0.021	-0.74
Log prop. in ward w/o qualifications x Female	-0.054	-0.84	-0.059	-0.98		
Log of practice list size in `000s	0.129	5.36	0.117	5.97	0.117	7.12
Log of practice list size x Female	-0.021	-0.48	-0.007	-0.18		
Log of practice list size per WTE GP	0.021	0.37	0.086	1.95	0.086	2.24
Log of list size per WTE GP x Female	-0.052	-0.53	0.019	0.22		
Log of age/sex adjustment	-1.101	-3.64	-0.695	-2.61	-0.684	-3.03
Log of age/sex adjustment x Female	0.105	0.18	-0.108	-0.22		
Log of nursing home adjustment	0.002	0.45	0.014	2.14	0.011	2.56
Log of nursing home adjustment x Female	0.006	0.81	-0.006	-0.69		
Log of additional needs adjustment	0.262	1.31	0.095	0.59	0.193	1.36
Log of additional needs adjustment x Female	-0.011	-0.03	0.209	0.66		
Dispensing practice	0.138	5.37	0.155	7.33	0.159	8.07
Dispensing practice x Female	0.018	0.36	0.014	0.30		
PMS practice	0.090	4.10	0.086	5.06	0.074	4.86
PMS practice x Female	-0.005	-0.13	-0.037	-1.04		
Ex-fundholder practice	0.069	3.18	0.054	3.31	0.035	2.48
Ex-fundholder practice x Female	-0.053	-1.41	-0.039	-1.18		
GOR1 (ref.cat.)						
GOR2	0.023	0.48	0.002	0.04	0.054	1.33
GOR2 x Female	0.114	1.21	0.138	1.50		
GOR3	0.049	1.00	0.045	1.01	0.106	2.52
GOR3 x Female	0.157	1.65	0.163	1.69		
GOR4	0.059	1.10	0.051	1.11	0.085	1.92
GOR4 x Female	0.031	0.30	0.071	0.69		
GOR5	0.063	1.32	0.051	1.21	0.092	2.19
GOR5 x Female	0.107	1.03	0.097	0.96		
GOR6	0.081	1.54	0.081	1.79	0.113	2.55
GOR6 x Female	0.045	0.42	0.076	0.73		
GOR7	0.062	1.14	0.054	1.07	0.101	2.17
GOR7 x Female	0.091	0.82	0.121	1.16		
GOR8	0.016	0.32	0.041	0.87	0.080	1.83
GOR8 x Female	0.118	1.18	0.102	1.04		
GOR9	0.075	1.37	0.062	1.36	0.089	2.10
GOR9 x Female	0.054	0.55	0.066	0.68		
Constant	5.487	4.13	4.713	4.07	4.136	4.08
R-squared	0.131		0.463		0.455	
N	1776		1776		1776	

Table 3. Predicted wages and income versus hours worked by GP gender

Hours	N female GPs	Wage female GPs	Income female GPs	N male GPs	Wage male GPs	Income male GPs	N all GPs	Wage all GPs	Income all GPs	Proportion female GPs
less than	22	43.36	30.82	13	109.35	53.67	35	67.87	39.31	0.63
20-24	54	40.30	42.13	21	68.72	71.87	75	48.26	50.46	0.72
25-29	97	40.01	49.32	21	59.71	74.06	118	43.51	53.72	0.82
30-34	107	35.69	51.85	34	52.69	76.09	141	39.79	57.69	0.76
35-39	69	35.01	59.63	71	45.85	78.29	140	40.51	69.10	0.49
40-44	87	32.85	63.24	197	41.20	79.57	284	38.64	74.57	0.31
45-49	56	30.82	66.56	218	37.43	81.50	274	36.08	78.44	0.20
50-54	58	28.50	67.82	269	34.97	82.87	327	33.82	80.20	0.18
55-59	13	26.32	68.82	113	32.06	83.50	126	31.47	81.99	0.10
60-64	26	27.66	78.74	144	29.65	83.98	170	29.35	83.18	0.15
more	9	25.61	85.45	77	25.70	85.54	86	25.69	85.53	0.10

Income is £000s per year. Wage is £s per hour = income/(weekly hours*47)

Table 4. Decompositions based on pooled log wage and income regressions with full sets of gender interactions.

	Decompositions with male coefficients as counterfactual							
	Wage model				Income model			
	Explained		Unexplained		Explained		Unexplained	
Characteristics	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**
Log hours					0.074	4.66	-1.139	-3.67
Personal characteristics	0.007	1.21	0.267	1.59	0.012	2.61	0.310	2.12
Practice characteristics	0.006	1.31	-0.356	-0.13	0.021	3.68	-0.766	-0.33
Geographic characteristics	0.002	0.96	-0.011	-1.40	0.001	0.77	-0.010	-1.29
Total excluding constant	0.015	2.16	-0.100	-0.04	0.108	6.07	-1.605	-0.70
Constant			0.167	0.06			1.881	0.82
Total including constant			0.068	3.60			0.275	10.78
	Decompositions with female coefficients as counterfactual							
	Wage model				**Income model**			
Characteristics	**Explained**		**Unexplained**		**Explained**		**Unexplained**	
	Coef.	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**
Log hours					0.171	8.15	-1.236	-3.67
Personal characteristics	0.025	1.96	0.250	1.48	0.042	3.79	0.280	1.92
Practice characteristics	0.005	0.86	-0.355	-0.13	0.011	1.98	-0.756	-0.32
Geographic characteristics	-0.003	-0.80	-0.007	-0.92	-0.002	-0.83	-0.006	-0.88
Total excluding constant	0.027	2.06	-0.111	-0.04	0.221	10.01	-1.718	-0.75
Constant			0.167	0.06			1.881	0.82
Total including constant			0.056	2.46			0.162	6.69

Table 5. Decompositions of gender differences in mean log income from propensity score matching

	Stratification		Nearest neighbour		Radius		Kernel	
	Coef.	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**
d_o	0.237	6.89	0.211	4.96	0.266	8.34	0.250	8.26
d_x	0.147	4.85	0.096	2.77	0.014	1.07	0.134	4.95
d_M	0.000	0.03	0.076	3.38	0.024	1.67	0.000	0.01
d_F	.	.	.	.	0.079	3.41	.	.
Total	0.384	22.58	0.384	22.58	0.384	22.58	0.384	22.58

Table 6. Decompositions of gender differences in mean income from propensity score matching

	Stratification		Nearest neighbour		Radius		Kernel	
	Coef.	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**	**Coef.**	**Bootstrap t-stat.**
d_0	15.521	9.00	13.788	5.76	16.721	8.00	16.040	9.66
d_x	8.425	5.75	5.259	2.75	1.035	1.08	7.906	5.32
d_M	0.031	0.31	4.930	3.17	1.958	1.84	0.031	0.14
d_F	.	.	.	.	4.263	3.29	.	.
Total	23.976	23.01	23.976	23.01	23.976	23.01	23.976	23.01

Table 7. Test for discrimination: log income model with polynomials in list size and average list size per GP

	Full set of interactions		Reduced set of interactions	
Variable	**Coef.**	**t-stat.**	**Coef.**	**t-stat.**
Unisex practice (U)	0.892	2.07	0.029	1.07
Unisex practice x Female	0.036	0.04	0.046	0.67
Female (F)	-1.232	-3.09	-1.218	-3.51
Log of weekly hours worked	0.293	4.54	0.254	4.72
Log of weekly hours worked x F	0.332	3.42	0.323	3.79
Log of weekly hours worked x U	-0.218	-2.15		
Log of weekly hours worked x F x U	-0.198	-1.23		
Experience in decades	0.094	0.97	0.095	1.14
Experience x F	-0.278	-1.73	-0.287	-1.98
Experience x U	-0.002	-0.01		
Experience x F x U	0.484	0.92		
Experience squared	-0.023	-1.06	-0.025	-1.33
Experience squared x F	0.072	1.89	0.077	2.27
Experience squared x U	-0.004	-0.12		
Experience squared x F x U	-0.080	-0.74		
Non-white	-0.043	-1.39	-0.019	-0.68
Non-white x F	0.186	3.34	0.179	3.53
Non-white x U	0.072	1.40		
Non-white x F x U	0.059	0.41		
Non-UK qualified	-0.011	-0.45	-0.007	-0.29
Salaried GP	-0.083	-1.94	-0.083	-1.93
Senior GP	0.081	4.43	0.081	4.46
Log of LISI score	-0.064	-3.12	-0.062	-3.00
Log of prop. in ward w/o qualifications	-0.019	-0.67	-0.018	-0.64
Log of age/sex adjustment	-0.644	-2.89	-0.688	-3.03
Log of nursing home adjustment	0.011	2.61	0.010	2.51
Log of additional needs adjustment	0.214	1.52	0.179	1.26
Dispensing practice	0.164	8.30	0.162	8.19
PMS practice	0.072	4.68	0.069	4.49
Ex-fundholder practice	0.031	2.17	0.033	2.34
Log of practice list size in `000s	0.029	1.08	0.029	1.08
Log of practice list size in `000s ^2	0.000	0.03	0.000	0.02
Log of practice list size in `000s ^3	0.000	-0.59	0.000	-0.56
Log of practice list size in `000s ^4	0.000	1.04	0.000	1.00
Log of practice list size per WTE GP	-1.361	-1.72	-1.190	-1.47
Log of practice list size per WTE GP ^2	0.802	1.84	0.699	1.57
Log of practice list size per WTE GP ^3	-0.188	-1.87	-0.162	-1.57
Log of practice list size per WTE GP ^4	0.015	1.87	0.013	1.55
GOR2	0.057	1.42	0.051	1.25
GOR3	0.111	2.63	0.104	2.46
GOR4	0.067	1.53	0.062	1.39
GOR5	0.096	2.27	0.089	2.09
GOR6	0.117	2.66	0.108	2.43
GOR7	0.106	2.27	0.098	2.08
GOR8	0.081	1.87	0.075	1.71
GOR8	0.097	2.27	0.090	2.09
Constant	4.630	4.01	4.988	4.25
R-squared	0.471		0.462	
N	1776		1776	

Table 8. Differences in mean income by gender mix of practice

Difference in conditional means	Coef.	t-stat.
$y_{ff} - y_{mm} = \alpha_0 + \delta_0 + \bar{x}_1^F \alpha_1$	-0.232	-3.27
$y_{ff} - y_{fm} = \gamma_0 + \delta_0$	0.075	1.10
$y_{mm} - y_{mf} = \gamma_0$	0.029	1.07
$y_{fm} - y_{mf} = \alpha_0 + \bar{x}_1^F \alpha_1$	-0.278	-11.03
$(y_{fm} - y_{mf}) - (y_{ff} - y_{mm}) = -\delta_0$	-0.046	-0.67
$y_{ff} - y_{mm} = \alpha_0 + \delta_0 + \bar{x}_1^F (\alpha_1 + \delta_1)$	-0.313	-3.28
$y_{ff} - y_{fm} = \gamma_0 + \delta_0 + \bar{x}_1^F (\gamma_1 + \delta_1)$	0.054	0.59
$y_{mm} - y_{mf} = \gamma_0 + \bar{x}_1^M \gamma_1$	0.033	1.14
$y_{fm} - y_{mf} = \alpha_0 + \bar{x}_1^F \alpha_1$	-0.266	-9.51
$(y_{fm} - y_{mf}) - (y_{ff} - y_{mm}) = -(\delta_0 + \bar{x}_1^F \delta_1)$	0.047	0.48

y_{ff} (y_{mm}) mean income of female (male) GP in single sex practice conditional on covariates; y_{fm} (y_{mf}) conditional mean income of female (male) GP in mixed sex practice. $\bar{x}_1^F$ and $\bar{x}_1^M$ are vectors of the means of log hours, experience, experience squared and ethnicity for female and male GPs, respectively.

Figure 1. Predicted income versus hours worked by GP gender

Figure 2. Predicted wages versus hours worked by GP gender

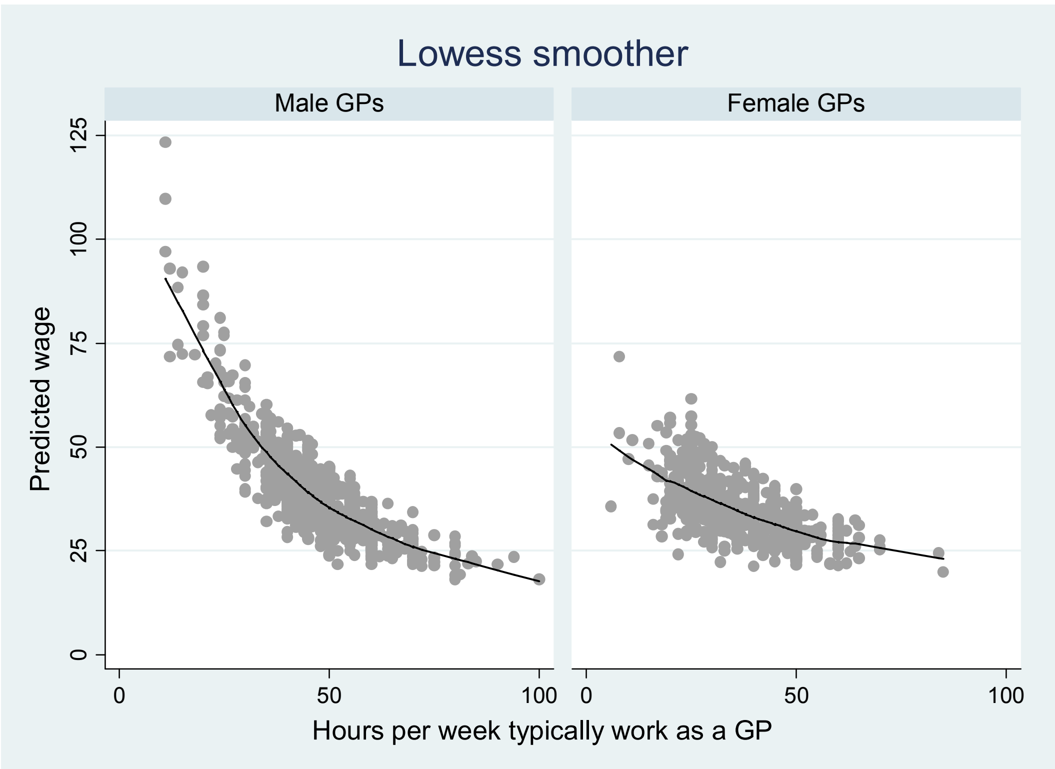
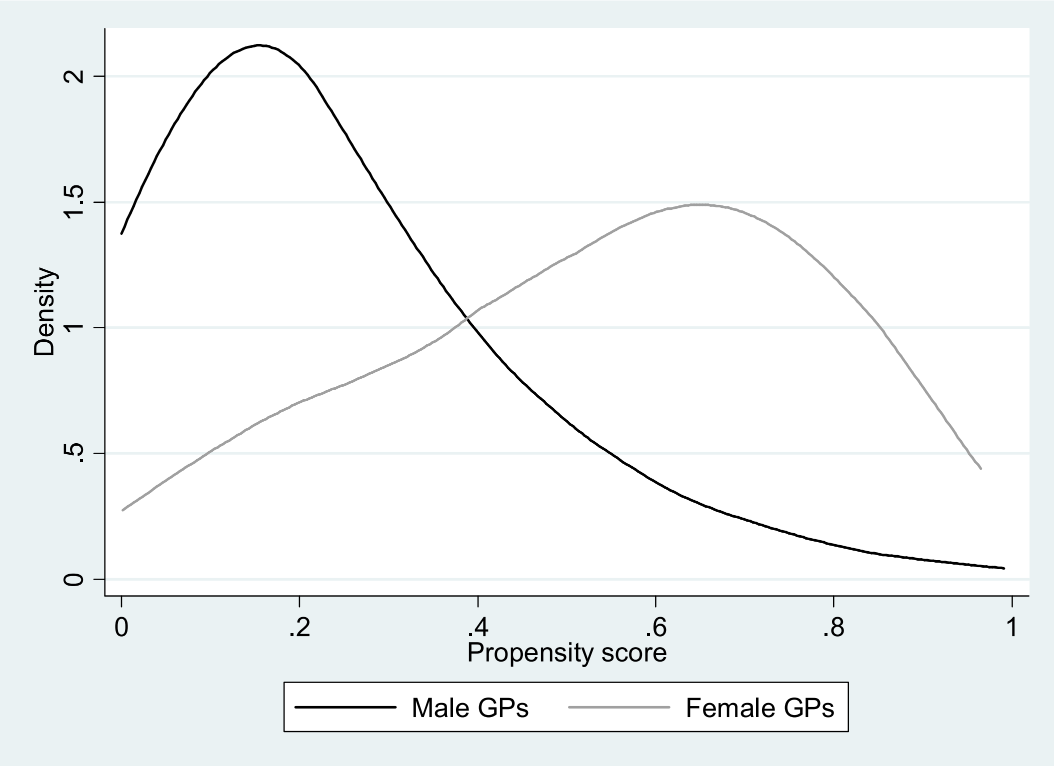


Figure 3. Kernel density plots of propensity scores by GP gender



Appendix Table 1. Propensity score model (probit regression with female as dependent variable).

Variable	Coef.	t-stat.
Weekly hours worked	-0.046	-14.78
Married - partner does not work	-0.562	-5.25
No children (ref.cat.)		
One child under 18	-0.385	-3.30
Two children under 18	-0.462	-4.49
Three or more children under 18	-0.718	-6.09
Experience in decades	-0.377	-1.32
Experience squared	0.018	0.28
Non-white	-0.246	-1.88
Non-UK qualified	-0.026	-0.18
Salaried GP	-0.499	-1.99
Senior GP	-0.564	-5.76
Log of LISI score	-0.072	-0.76
Log of prop. in ward w/o qualifications	-0.050	-0.40
Log of practice list size in `000s	-0.485	-5.26
Log of practice list size per WTE GP	0.540	2.65
Log of age/sex adjustment	-0.425	-0.38
Log of nursing home adjustment	-0.145	-4.52
Log of additional needs adjustment	0.537	0.89
Dispensing practice	-0.068	-0.68
PMS practice	-0.060	-0.78
Unisex practice	-0.599	-4.30
Ex-fundholding practice	-0.045	-0.60
Constant	2.992	0.63
Pseudo R-squared	0.259	
N	1776	