THE UNIVERSITY *of York*

*Discussion Papers in Economics*

No. 11/05

# Measuring and testing for gender discrimination in physician pay: English family doctors

By

Hugh Gravelle   Arne Risa Hole   Rita Santos

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

# Measuring and testing for gender discrimination in physician pay: English family doctors

**Hugh Gravelle*  Arne Risa Hole**  Rita Santos***

**Abstract**

In 2008 the income of female GPs was 70%, and their wages (income per hour) were 89%, of those of male GPs. We estimate Oaxaca decompositions using OLS models of wages (income/hours) and 2SLS models of income. The elasticity of income with respect to hours is 0.91 for female GPs and 0.29 for male GPs, so that log wage models are misspecified. The conventional discrimination measure (the unexplained difference in mean log income) is sensitive to the counterfactual (30% using male returns vs 11% using female returns), to the use of OLS vs. 2SLS (19% vs 11%, female counterfactual), but not to dropping insignificant female interactions. The unexplained pro-male difference arises because the pro male difference in regression constants offsets the pro-female difference in the effect of hours on income.

We propose a set of new direct tests for within workplace gender discrimination based on a comparison of the differences in income of female and male GPs in practices with varying proportions of female GPs and with female or male senior partners. The direct tests produce mixed results. An indirect test, comparing GPs actual income with the income they report as an acceptable reward for their job, shows that female GPs are not more likely than male GPs to report that their actual income is less than acceptable income, whereas GPs from ethnic minorities and overseas qualified GPs are significantly more likely to do so.

JEL Nos: J16, J44, J71, I11

Keywords: Gender discrimination. Family doctors. General practitioners. Income. Wages.

---

* National Primary Care Research and Development Centre, Centre for Health Economics, University of York; York YO10 5DD. Corresponding author: Gravelle: email: hg8@york.ac.uk; phone 01904 321418; Fax 01904 321402.

**Department of Economics, University of Sheffield.

# 1 Introduction

## 1.1 Gender discrimination in physician pay

In this paper we analyse the substantial differences in income between male and female general practitioners (GPs) in the English National Health Service. In 2008 the income of female GPs was 69%, and their wages (income per hour) were 91%, of those of male GPs. There are similar gender differences in physician pay in the US and elsewhere.

Our study makes a number of contributions. First, it is the only study of gender income differentials in English GPs who account for the majority of patient contacts with the NHS and who act as gatekeepers to NHS inpatient and outpatient care. By restricting attention to a group 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.

Our second contribution is to explore the robustness of conventional estimates of discrimination between male and female GPs to alternative ways of dealing with a number of methodological issues. In the standard approach to measurement of discrimination, models of pay are estimated for each gender and the difference in mean pay is decomposed into an "unexplained" component due to differences in estimated returns to worker characteristics and an explained component due to differences in characteristics (Blinder, 1973; Oaxaca, 1973). The unexplained component is conventionally taken as a measure of discrimination.

We find, as in the study of US physicians by Bashaw and Heywood (2001), that the size of the unexplained component is much larger for income than wages measured as income divided by hours (30% vs 6% assuming female GPs would be rewarded in the same way as males in the absence of discrimination). The reason is that the elasticity of male GP income with respect to hours is only 0.28 compared with 0.91 for female GPs. Hence the measured "wage" (income/hours) declines with hours worked, and since male GPs work longer hours than female GPs, the difference in wages is smaller than the difference in incomes.

Models of physician gender pay discrimination should therefore be based on models of the determinants of income (including hours worked), rather than models of wages. But, of previous studies of physician income gender differences, only Conrad et al (2001) has considered the possibility that hours may be endogenous so that the usual OLS regression models of income yield biased coefficients. We find that female GP hours are endogenous and so, unlike previous studies, we estimate a two stage least squares model of income. If endogeneity of hours is not allowed for then the unexplained component of the log income difference is nearly doubled (19% vs 11%, assuming that all GPs would be rewarded at the female rate in absence of discrimination).

The unexplained component in the conventional decomposition of income differences includes the difference in regression constants. This is problematic since the estimated constants depend on the effects of unobserved characteristics on pay and on the average value of the unobserved characteristics. Thus the difference in the regression constants from the male and female income models could be entirely due to differences in returns or entirely due to differences in the joint distribution of the unobserved and observed characteristics. Thus the standard two way decomposition is potentially misleading. We therefore suggest reporting a three way decomposition into parts due to the difference in constants, the difference in returns on observed variables, and the difference in the means of the observed variables. We find that the unexplained pro-male difference in the standard decomposition arises because the pro male difference in regression constants offsets the pro-female difference in the effect of hours on income.

The third contribution of the paper is to propose direct and indirect methods of testing for pay discrimination. English family doctors typically work in small partnerships of 3 to 4 GPs and undertake a range of activities with different effects on practice profit and patient health. The income of a GP depends on the total profit of their practice and the within practice profit sharing arrangements. The within practice profit sharing rules can lead to gender differences in pay in three ways. Female GPs may be paid less for any mix of activities; this is discrimination. Female GPs may also have lower incomes because they undertake a mix of activities which generates a lower share of the practice profit. Whether this is discrimination depends on whether female GPs choose their less profitable activity mix or whether it is assigned to them.

Our first direct test for within practice discrimination rests on the fact that there cannot be within practice gender discrimination in practices where all the GPs have the same gender. Thus the difference in income between female and male GPs in otherwise similar all female and all male practices is due to differences in productivity or in preferences. The unexplained difference between female and male GP incomes in otherwise similar mixed gender practices ($y_{fm} - y_{mf}$), minus the difference between female and male incomes in single gender practices ($y_{ff} - y_{mm}$), is an estimate of the extent of within practice gender discrimination. We find that this difference [($y_{fm} - y_{mf}$) − ($y_{ff} - y_{mm}$)] is pro-male for four specifications of the GP income model, with estimates of discrimination varying between 31% (p=0.113) to 12% (p=0.635) and being significant in one of the specifications (29%, p=0.005).

The direct test using the incomes in single gender practices is vulnerable to two difficulties: There are relatively few female GPs in all female practices (35 of 753 female GPs) so that, although males in all male practices have larger incomes than females in all female practices, the difference ($y_{ff} - y_{mm}$) is very imprecisely estimated. The second difficulty is that if there is discrimination in the selection of female GPs into practices which are potentially less profitable even after conditioning on observed practice characteristics, then the difference ($y_{ff} - y_{mm}$) will overstate the difference in productivities.

Our second and third direct tests of gender discrimination are less vulnerable to these difficulties. They rest on the plausible assumption that female GPs are less likely to discriminate against other female GPs. If there is discrimination against female GPs then being in a practice with a higher proportion of female partners or with a female senior GP should increase the income of female GPs relative to male GPs. We find that the gender of the senior partner has no effect on the income of male or female GPs and that a higher proportion of female GPs tends to reduce the income of both male and female GPs. The latter finding suggests that, even if there is selection of female GPs into less lucrative practices, there is no gender discrimination in profit sharing.

Our indirect tests for gender discrimination use GP reports of overall job satisfaction, satisfaction with remuneration, and their estimates of the income they consider to be adequate for their job given their experience and career stage. We find that female GPs are if anything more satisfied with their job overall and with their pay, conditional on their hours and income. Although female GPs report that an adequate income for their job exceeds their actual income, the difference relative to male GPs is small and not significant. By contrast non-white and overseas qualified GPs both report adequate incomes for their job which are much larger relative to white and UK qualified GPs.

The results from our tests for discrimination provide only weak support for the existence of within practice discrimination. Our finding that GPs in practices with a higher proportion of female GPs have lower incomes could be due to selection of female GPs into less profitable practices but could also be due to gender differences in preferences or productivity across activity mixes. If there is gender discrimination it is surprising that it does not manifest itself in dissatisfaction with pay or in a larger difference between reported adequate and actual income for female GPs.

In the second part of this introduction we briefly summarise the relevant literature on gender differences in physician pay. Section 2 discusses methodological questions in the measurement of gender pay discrimination. Section 3 provides background information on the NHS and the contracts under which practices are paid. Section 4 describes the data and section 5 presents the decomposition results. Section 6 describes our proposed tests for gender discrimination and the results of the tests are in presented in section 7. Section 8 has a brief summary and conclusion.

## 1.2 Related literature

There are gender differences in pay even for individuals with considerable investments in human capital working in the same profession (Wood et al, 1993; Dolton et al, 1996; Blackaby et al, 2005; Noonan et al, 2005; McNabb and Wass, 2006; Bertrand et al, 2009). In medicine marked gender differences in income and hourly pay are reported in many countries (Robinson, 1998; Gupta et al, 2003; Theurl and Winner, 2010).

Almost all previous studies of gender differences in physician pay are for the US. Kehrer (1976) found a 30% overall difference and a 22% unexplained difference in hourly wages in 1973. Langwell (1982) found a 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 estimated 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 and in primary care female physicians had 13% higher hourly wages. Bashaw and Heywood (2001) showed 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 reported a 17%

unexplained difference in income. Ash et al (2004) found that female US medical academics are less likely to hold higher rank and have lower pay given their rank. Their pay was 13% less than if they were rewarded for their characteristics at the same rate as men. Sasser (2005) investigated the effect of family circumstances on the earnings of young (under 40) US physicians in 1986 an 1990. Being married reduced female earnings relative to male physicians by 11% and being married with more than one child reduced them by a further 22%. It was not the case that less productive female physicians were more likely to get married and to have children. Shih and Konrad (2007) report that in 1996/7 male primary care physicians had log incomes 23% higher and log wages 14% higher than female primary care physicans after controlling for a large range of factors. The only non-US study reports a 32% pro male total difference in log income and a 15% unexplained difference for Austrian physicians in 2002-4 in a public insurance system in which fees are controlled by the insurance fund (Theurl and Winner, 2010).

# 2 Measuring gender discrimination

## 2.1 Regression decomposition

The standard method for analysing the difference in pay between male and female workers is to estimate separate Mincer (1970; 1974) remuneration models for males and females: $\ln w_i^k = \beta_0^k + \mathbf{x}_i'^k \boldsymbol{\beta}^k + \varepsilon_i^k$, where the superscripts $k$ = $M$, $F$ denote gender. The difference in mean log wages is decomposed into an unexplained part due to differences in the estimated coefficients and an explained part due to differences in the means of the explanatory variables (Blinder, 1973; Oaxaca, 1973):[1]

$$\overline{\ln w^M} - \overline{\ln w^F} = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^F) + \overline{\mathbf{x}}^{F\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)}_{\text{Unexplained}} + \underbrace{(\overline{\mathbf{x}}^{M\prime} - \overline{\mathbf{x}}^{F\prime})\hat{\boldsymbol{\beta}}^M}_{\text{Explained}} \quad (1)$$

The percentage 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 observed male rate. Medicine has traditionally been a male dominated profession which would support this assumption and we adopt it in most of calculations. However, the proportion of female GPs in England is now over 40%, and to illustrate the sensitivity of the estimate of discrimination to the assumed counterfactual, we also present calculations using female coefficients as the counterfactual.

*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 and hours

---

[1] Recently propensity score matching methods have also been used to measure explained and unexplained differences in mean incomes between different groups (Barsky et al, 2002). In an earlier work we found that the decompositions from propensity score matching were very similar to those from regressions (Gravelle and Hole, 2008).

can be dropped from the pay model. But if pay is not proportional to hours, a pay 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.

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 et al, 1993), we focus on differences in income and decompositions derived from earnings equations:

$$\overline{\ln y^M} - \overline{\ln y^F} = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^F) + \overline{\mathbf{x}}^{F\prime}(\hat{\boldsymbol{\beta}}^{xM} - \hat{\boldsymbol{\beta}}^{xF}) + \overline{\ln h^F}(\hat{\beta}^{hM} - \hat{\beta}^{hF})}_{\text{Unexplained}}$$
$$+ \underbrace{(\overline{\mathbf{x}}^{M\prime} - \overline{\mathbf{x}}^{F\prime})\hat{\boldsymbol{\beta}}^{xM} + (\overline{\ln h^M} - \overline{\ln h^F})\hat{\beta}^{hM}}_{\text{Explained}} \qquad (2)$$

where $y$ is earnings and $h$ is hours.

*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 other variables in the income regression will also be biased if they are correlated with hours worked.

In this paper, unlike the previous literature which has examined differences in income (Bashaw and Heywood, 2001; McNabb and Wass, 2006; Theurl and Winner, 2010; Wood et al, 1993), we test for and correct potential endogeneity bias by using two stage least squares (2SLS) rather than OLS.

*Regression constants*.
In the conventional decomposition, differences in the regression constants are counted as part of the unexplained difference. This is problematic. The constant in a linear 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. Thus the gender difference in constant terms conflates differences in the unobserved coefficients on the omitted variables (which would be counted in the "unexplained" component if the omitted variables were observed) and the differences in the joint distributions of the omitted

and included variables (which would be counted in the "explained" part if the omitted variables were observed). [2]

We provide a three way decomposition based on the regression model: labelling the difference in constants as "unknown" and relabeling the other two parts as due to differences in returns for observed characteristics and differences in observed characteristics:

$$\overline{\ln y^M} - \overline{\ln y^F} = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^F)}_{\text{Unknown}} + \underbrace{\overline{\mathbf{x}}^{F\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F) + \overline{\ln h^F}(\hat{\beta}^{hM} - \hat{\beta}^{hF})}_{\text{Difference in returns}}$$
$$+ \underbrace{(\overline{\mathbf{x}}^{M\prime} - \overline{\mathbf{x}}^{F\prime})\hat{\boldsymbol{\beta}}^{xM} + (\overline{\ln h^M} - \overline{\ln h^F})\hat{\beta}^{hM}}_{\text{Difference in characteristics}} \qquad (3)$$

Because we use the Yun (2005) procedure to allow for arbitrariness of the omitted baseline categories of the categorical variables in **x,** the "unknown" part of the decomposition will be an average of the difference in the constants over all possible choices of baseline categories. However, our point still applies: the difference between the regression constants for males and females can arise from unobserved differences in returns and from differences in the means of unobserved variables and from differences in the returns on these omitted variables. It is arbitrary to add it to either of the other two categories.

*Insignificant gender differences in coefficients*

The discrimination measure depends on the difference between male and female 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 of estimating a model with a full set of gender interactions on all explanatories, including the insignificant ones, and then to calculate the standard error for the "unexplained" component of the difference in mean pay including insignificantly different male and female coefficients.

The second method, which is less usual in the literature, is to estimate a model with a restricted set of gender interactions, as in Wood et al (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 decompositions based on models with full and restricted sets of gender interactions.

[2] The partition of the unexplained component between the difference in constants and the difference in returns is also not invariant to the addition or subtraction of a constant from the explanatory variables (Jones, 1983; Oaxaca and Ranson, 1999). For example, defining experience as age minus 25, rather than age or age minus 27, does not affect the coefficients on experience but does affect the constant terms. We have a natural origin of zero for our experience variable since it is measured as reckonable service which is akin to full time equivalent service in the NHS.

# 3 Family doctor contracts

The NHS is financed almost entirely from general taxation and patients face no charges for NHS health care, apart from a small charge for dispensed medicines.[3] Patients register with general practices, which act as gatekeepers for elective hospital care. Practices have on average 4.1 GPs and around 6,500 patients. In 2008 43% of GPs were women (Information Centre, 2009b).

GPs are not employees of the NHS, apart from a small proportion directly employed by local primary care organisations (Primary Care Trusts - PCTs). The NHS contracts with general practices, not with individual GPs. Nearly all practices are legal limited liability partnerships of a small number of GPs. The GPs in the practices are usually profit sharing partners (92% in our sample) rather than salaried employees of the practice.

Practice contracts with the NHS to supply services to patients are of two types. Just over half of general practices have the General Medical Services (GMS) contract whose terms are set by national negotiations between the NHS and the British Medical Association (the doctors' trade union). GMS practices are paid a mixture of capitation, lump sums, items of service, and quality incentive payments. Most of practice revenue (over 60%) is generated by capitation payments which are determined by a national formula which takes account of the demographic mix of practice patients and local morbidity measures. Quality incentives from the Quality and Outcomes Framework (Roland, 2004) generate over 20% of practice revenue. The remainder of practice revenue consists of lump sum payments (related to the location of the practice, newly qualified GPs, and seniority payments), fee per item payments (mainly for vaccinations), and payments for vaccinating and screening specified target proportions of the relevant practice population. GPs are also permitted to dispense the medicines they prescribe to patients without easy access to a pharmacy. 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. Practices are reimbursed for the costs of their premises but have to fund all other expenses, such as hiring practice nurses and clerical staff, from their revenue.

Around 45% of practices are 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. PMS practices also receive QOF payments, though they are paid less than GMS practices for the same quality achievement because some of the QOF payments relate to activities which are also paid for directly under PMS contracts. As under GMS, the practice has to meet its expenses from its revenue.

---

[3] 60% of patients are exempt from charges and 90% of prescriptions are dispensed without charge in 2008 (Information Centre, 2009a).

# 4 Data

The data are from the 2008 National Primary Care Research and Development Centre's General Practitioner Worklife survey of English GPs (Hann et al, 2009). We use responses from 734 female and 1168 male GPs. GPs were asked to report their banded[4] earnings, defined as total annual income from their practice before taxes but after deducting expenses from their practice; their usual hours worked per week and the number of weeks holiday per year; and their personal and practice characteristics.

We added further information regarding the GPs' practices from the General Medical Services data set. Data on local population socio-economic characteristics from the Office of National Statistics Neighbourhood Statistics data set were attributed to practices using information on the shares of practice populations in each Lower Super Output Area[5] from the Attribution Data Set and the General Medical Services data set, both maintained by the Department of Health.

We divide the variables potentially affecting GP remuneration into:

*Personal characteristics*.
We measure hours worked as annual hours worked, calculated as usual weekly hours multiplied by 52 minus the number of weeks holiday.[6] We measure experience as decades of reckonable service. Reckonable service is used to calculate GP pensions and is the full-time equivalent years the GP has spent working for the NHS. It is a better measure of experience than age or years since qualification which make no allowance for part-time working or years out of the workforce.[7]

We include self reported 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. Because we also include ethnicity the coefficient on non-UK qualified may reflect the impact of less relevant initial experience on earnings.

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, whether their partner works full time or part time or not at all, the number of children under 18 and the age of the youngest child.

---

[4] The earnings bands (% of GPs in each band) were: less than £25,000 (1.22%); £25,000-£49,999 (8.11%); £50,000-£74,999 (16.17%); £75,000-£99,999 (26.01%); £100,000-£124,999 (29.22%); £125,000-£149,999 (12.49%); £150,000-£174,999 (3.82%) ; £175,000 or more (2.96%). For the log income regressions we use the mid points of the logarithms of the income bands. The income data were reasonably well described by a normal distribution (mean £97,703, variance £1,248,087) truncated at £0 and £200,000. For the open ended upper band we use the logarithm of £184,970 which is the predicted mean of this band from the normal distribution. Results were not sensitive to alternative assumptions about the mean of the open ended upper band nor to the use of interval regressions (Jones, 2000).

[5] Lower Super Output areas have minimum populations of 1000 people and a mean population of 1500.

[6] Models with log weekly hours and log weeks worked entered separately yielded insignificant coefficients on log weeks worked (perhaps because there was little variation in this variable) and coefficients on log weekly hours which were similar to those from the models with log annual hours.

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

We include an indicator for whether the GP is the senior partner in the practice. We expect that senior partners will have higher incomes than otherwise similar GPs since their role has more responsibility. We also include an indicator for salaried status and expect that it will have a negative effect on earnings.

*Practice characteristics*.

Capitation payments, which account for over 60% of practice revenue, are proportional to the number of patients on the practice list, multiplied by four indices reflecting local population morbidity (additional needs index), the demographic mix of the patients (age/sex index), the proportion of patients in nursing homes, and the turnover of the practice list.[8] We have measures of the first three of these and include them in the regression models. Practice demographics, the morbidity of practice patients, and the proportion in nursing homes will also affect the costs of practices so that it is unclear a priori if these variables will be positively or negatively associated with the income of GPs which depends on the difference between practice revenue and costs.

We also include the total practice list size and the number of patients per GP. Capitation payments will increase pro rata with the practice list but treating more patients requires more inputs and so higher practice costs. 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 because monitoring problems lead to attenuation of incentives for income generating effort. US evidence suggests that there are no strong economies of scale in physician practice (Pope and Burge, 1996) but little is known about cost functions in UK general practice. Thus the effect of list size on GP income could be negative or positive. We also have indices of the cost of inputs (staff, premises) which vary geographically.

Over 20% of practice revenue is generated by the Quality and Outcomes Framework which rewards practices for achieving a range of quality indicators. We do not have data on QOF revenue but QOF payments are well approximated by the number of points earned for achieving indicators multiplied by the practice list. We therefore include the number of QOF points.

Practices can earn additional income by providing fee per item services, such flu vaccinations, by meeting targets for vaccination of children and cervical screening, and by providing additional services outside the NHS. 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 et al, 1995).

We have information on whether practices have opted to have extended opening hours or to provide out of hours cover to their patients, both of which would produce additional payments to the practice.

[8] The formula for calculating the capitation payments changed in 2004 but, because the new formula would have led to large reductions in practice revenue, 90% of practices receive a capitation payment based on the formula in use up to 2004 (National Audit Office, 2008). The variables we use are those from the pre-2004 formula.

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.

*Area measures*.
We have a measure of the rurality of the area in which the practice is located and measures, reported by GPs, of local amenties, housing costs, job opportunities for spouses, and the quality of local schools. We also include the number of other practices within 1km of the GP's practice as a measure of competition for patients. We also use 8 Government Office Region dummies to capture unobserved factors with geographically varying means.

# 5 Measuring gender discrimination

## 5.1 Descriptive statistics

Table 1 has descriptive statistics for the estimation sample. The mean annual gross earnings for male respondents is £110,690 while the mean earnings among female respondents is £77,030 or about 70% of mean male earnings. Male GPs work 2107 hours per year (45.9 hours per normal week) and female GPs work 1581 hours per year (34.5 hours per normal week) on average, or about 75% of male annual hours. Female GPs have an hourly wage of £51.65 which is 89% of the male wage.

Male GPs have more reckonable service than female GPs (23.1 vs 20.4 years) and are more likely to be the Senior GP in a practice, to be a partner rather than salaried, non-white, and qualified outside the UK. They are also more likely to be married with a non-working partner and less likely to have children aged under 4 years.

Male and female GPs work in practices with similar characteristics in terms of list size, list per GP, the demographic and socio-economic mix of patients, and the type of practice contract.

The estimation sample has broadly similar characteristics to the GP population (Information Centre, 2009b) except that it has a smaller proportion of female GPs (39% female against 43% in the GP population). In particular there are no substantial differences in the proportions with PMS status (43% against 46%) and which are dispensing practices (17% against 18%).[9]

[9] Because the sample was of GPs rather than practices and larger practices have more GPs, the average list size (9570) and number of GPs (5.6) in respondents' practices is greater than the national practice means.

## 5.2 Estimation

The models for the decomposition analysis were first specified with a full set of female interactions

$$\ln z_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta}_1 + F_i\left(\alpha_0 + \mathbf{x}_i'\boldsymbol{\alpha}_1\right) + \varepsilon_i \qquad (4)$$

where $z_i$ was the hourly wage or annual earnings and in the log income model hours are included in $\mathbf{x}_i$. All models were estimated with heteroscedasticity robust standard errors and clustering by practice since the 1902 GPs were in 1583 practices.

We first estimated the log income model using all the potential explanatories shown in Table 1 both as main effects and interacted with female gender dummy *F*. We then dropped main effect variables (and their interaction with *F*) if the estimated coefficient was insignificant (using a cut off of p = 0.20) for male GPs, for the female interaction, and for the sum of the main and interaction effects (ie for female GPs). Otherwise the variable and its female interaction were retained in the model. The resulting reduced model with a full set of female interactions is reported as model 2 in Table 2. We then dropped all female interaction terms which were insignificant at 10% and re-estimated to produce the reduced model with a restricted set of female interactions reported as model 3 in Table 2. The interaction terms dropped for model 3 were jointly insignificant at the 5% level (p = 0.38). The Bayesian Information Criterion suggests that model 3 with a restricted set of interactions performs better than model 2.[10]

For the log wage we used the same variables as in the reduced log income model with a full set of female interactions and this is reported as model 1 in Table 2. We estimated the log wage model by OLS.

For the log income models we first ran 2SLS models separately for male and female GPs. We used four instruments for log hours: having a partner who works full time, whether the youngest child was aged 4 or less, youngest child between 4 and 18, and the interaction between having a full time working partner and youngest child under 4. The F statistic on the instruments in the first stage female GP hours model was 18.66 (p < 0.001), the model passed the over-identification test for instrument validity (p = 0.140), and female hours worked were endogenous (Durbin-Wu-Hausman test, p = 0.029). The instruments were unsatisfactory for male GPs hours (F = 1.20, p = 0.309) and we were unable to find any other instruments for male hours. The only other study of physician income that tests for endogeneity of hours is Conrad et al (2002) which found that neither male or female hours were endogenous.

We therefore report estimates from an OLS model for male GPs and a 2SLS model for female GPs. Note that even if the assumption that male hours is exogenous is incorrect, this does not affect the validity of the usual two fold decomposition of income differences into explained and unexplained components if female returns are the counterfactual.[11]

---

[10] BIC = -2 ln *L* + *k* ln (*n*), where *L* is the maximised likelihood of the estimated model, *k* is the number of explanatory variables plus 1, and *n* is the number of observations. A smaller BIC is better.

[11] If the 2SLS estimate of the female coefficients is consistent, then $(\overline{\mathbf{x}}'^M - \overline{\mathbf{x}}'^F)\hat{\boldsymbol{\beta}}^F$ is a consistent estimate of the explained component of the difference in mean log income. Hence $\overline{\mathbf{x}}'^M(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)$ is a consistent estimate of the remaining, unexplained, component. If male hours are endogenous, the male coefficients $\hat{\beta}_k^M$ on the individual

## 5.3 Regression results for decomposition

The results in Table 2 are generally in line with our prior expectations and the coefficients on variables which are common to the log wage and the log earnings models are qualitatively similar. The results of the log wage model are also broadly similar to those from another study which used the 2008 GP Worklife Survey (Morris et al, 2010) to investigate the differences between PMS and GMS practices.

*Hours worked and income*

The coefficients for log hours in the log earnings regressions in Table 2 are significantly less than one for male GPs suggesting that their earnings are not proportional to hours worked. A 1% increase in hours worked leads to a 0.28% increase in earnings for male GPs in model 3. The coefficient on female log hours is much larger than for male GPs (0.912) and is not significantly different from 1.

Studies of other professional groups also find that the marginal returns to hours worked declines with hours worked. For US physicians, 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. Other studies also find that male income is less responsive to hours than female income. Bashaw and Heywood (2001) estimate elasticities of 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. A possible explanation is that male GPs are expected to spend more years in a practice and so their pay can be related to their average hours input over a longer period, whereas female GPs will have their pay related to average hours over a shorter period.

Because the male wage declines faster with hours worked and male GPs work more 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, using income/hours as the measure of income earning opportunities for male GPs is invalid because it is determined by decisions on hours worked but hours are not included in the wage model. Thus gender differences in pay for GPs should be investigated with models of income, not wages.

*Other determinants of pay*

GPs with more than 10 years reckonable service earn more than less experienced GPs though the effect is smaller for female GPs. Being non-white reduces male GP income by 6% but may even increase female GP income.

Being in a dispensing practice or a PMS practice has a large positive and significant effect on wages and earnings. Salaried GPs have 50% lower wages than GPs who are profit sharing

explanatory variables $x_k$ will be inconsistent, so that a more detailed decomposition within the unexplained component is not warranted, but the overall decomposition with female counterfactuals is consistent.

partners. Since relatively few of the GPs in our sample are salaried (4% of male GPs, 14% of female GPs) salaried status accounts for little of the gross gender difference in income.[12]

The capitation payment formula adjustments for input prices and the proportion of patients from nursing homes had no effect on income. The formula adjustment for the age/sex mix reduces the income of female GPs whereas the adjustment for the additional needs of patients reduces the income of male GPs.

Being a senior partner has no direct effect on income when hours are controlled for in the income model but does affect hours worked, so that the higher unconditional income of senior partners is a reward for longer hours, rather than for bearing greater responsibility.

## 5.4 Decomposition results

Table 3 has the results of the decompositions of the differences in mean log wages and income from different models in Table 2 plus an additional OLS model of income. We report a threefold decomposition into an unknown part due to the difference in the constants, a part due to difference to returns to characteristics, and a part due to differences in characteristics.

*Log wage vs log income decompositions*.
The choice of the remuneration measure makes a considerable difference to the decomposition and hence to conclusions about the magnitude of discrimination as measured by the difference in returns. The log percentage difference in mean wages is 11% which is much smaller than the difference in mean log income of 45%.[13] The traditional unexplained log difference (difference in constants plus difference in returns) is 6% for wages and 30% for income. The difference in mean wages is smaller primarily because hours worked have very different effects on income for male and female GPs but also because their hours differ.

*Choice of counterfactual*
The choice of the counterfactual in the decomposition makes little difference to the log wage decomposition. However, it has a substantial impact on the log income decomposition. The standard unexplained component of the differences log income is smaller when the counterfactual is what female GPs would have earned if they had been rewarded as if they were male GPs: 11% vs 30%. The reason is the marginal reward for female hours is larger than the male marginal reward.

*"Unknown" contribution to "Unexplained" component*
The conventional decomposition of the overall difference into "unexplained" vs "explained" parts implicitly assumes that the differences in the regression constants are due to differences in returns on unobserved characteristics. But they could also be due to differences in the means of the unobserved characteristics. For GPs, the largest component in absolute value of the unexplained part is the difference in the regression constants. It has the opposite sign

[12] By contrast, in the US 1991 Survey of Young Physicians used by several studies of gender pay differences, 46% of male and 65% of female physicians are not owners or part owners of their practice (Bashaw and Heywood, 2001).
[13] Note that this exceeds both the difference in the log of the means (0.36) and the proportionate difference in mean incomes (0.30).

(positive i.e. pro-male) to the sum of the differences in returns to coefficients. The difference in returns component (ie excluding the difference in constants) is negative – indicating pro-female advantage – because the coefficient on hours is much larger for female GPs than for male GPs.

Thus if one includes the constant term there is a pro-male unexplained income difference but if the constant is not counted there is a pro-female unexplained income difference. Blinder (1973) in his study of racial differences in wages, notes that the total unexplained pro white component arose because of a large pro-white difference in constants which more than offset a pro-black difference in the other coefficients. Since the difference in the constants could be due to differences in returns or to differences in characteristics, it is questionable whether one can conclude that the conventional two fold decomposition of income differences into explained and unexplained components is even prima facie evidence of gender discrimination in GP income.

*Effect of dropping insignificant gender interactions*

The conventional decomposition is not affected greatly by the decision to estimate a model with a full set of female interactions (model 2) or to drop insignificant female interactions (model 3): the log difference which is unexplained (difference in constants plus the difference in returns) is approximately the same (30%) with a male GP counterfactual and decreases slightly, from 11% to 10% with a female GP counterfactual. The use of a full or restricted set of gender interactions does however make marked difference to the magnitude of the differences in returns. Thus with male coefficients as the counterfactual the log difference in returns increases by 0.58 and the difference in constants falls by 0.58.

*OLS vs 2SLS*

We found that female hours are endogenous in the model of income and so used 2SLS to estimate the income models reported in Table 2. The last two columns of Tables 3A and 3B report the decomposition based on an OLS income model with a full set of female interaction using the same set of variables as the 2SLS income model, but with actual rather than instrumented female hours.[14] The OLS decomposition using the male counterfactual is identical to the 2SLS decomposition because the estimated male coefficients are the same in these models since only female hours were endogenous. Table 3B uses female counterfactuals. With an OLS model the total unexplained difference is nearly twice as large as the unexplained difference in the 2SLS model: 19% log difference is unexplained as against 11% log difference. The difference in between the OLS and 2SLS model decompositions is due to the much smaller female hours coefficient in the OLS model (0.615 vs 0.906). Thus it is important to test and allow for endogeneity of hours when decomposing gender income differences.

# 6 Testing for within practice discrimination

We have shown in the previous sections that there are statistically and economically significant differences in pay between male and female GPs which are not explained by differences in their observable characteristics. In this section we discuss the possible reasons for these gender differences. Given total practice profit, GP pay is determined by within

[14] See Appendix Table A3

practice profit sharing arrangements. We suggest some direct and indirect tests of whether gender pay differences are due to within practice discrimination in the sharing of practice profits.

## 6.1 Direct tests for within practice gender discrimination

There are a number of possible reasons why female GPs earn less than otherwise observationally similar male GPs: (i) patient tastes for GP gender; (ii) selection of female GPs into less lucrative practices; (iii) within practice discrimination in the sharing of practice profit; (iv) preference or productivity differences between male and female GPs.

If patient demand for being on a practice list is affected by the gender mix in a practice then, in a healthcare system where practices set prices, practices with a more preferred gender mix may be able to charge higher prices (Reyes, 2006). However, there are no prices to patients for GP services in the NHS. It is possible that practices with greater demand would find it easier to engage in cream skimming to select a mix of patients for whom it is easier to achieve pay for performance targets or who impose less effort or other costs on the practice. If patients preferred male GPs to female GPs then female GPs would be paid less, ceteris paribus, to induce practices to take them on. But there is no evidence that male doctors are preferred to female doctors by patients in general. The gender of practice GPs has no effect on patient satisfaction with practices (Hann et al, 2008; Baker, 1995; Murphy-Cullen and Larsen, 1984). Consultations with female family physicians are longer and more patient centred than male physicians (Roter, et al, 2002). Thus patient preferences seem unlikely to explain lower female GP earnings.

It is possible that female GPs are disproportionately in practices with less lucrative patient lists. This might be because of discrimination against them or because they have preferences over other aspects of the practice (for example the distance to the practice from their home or the flexibility of working hours or the mix of patients on the practice list) and are willing to trade off income against other characteristics of practices. We have data on a rich set of practice characteristics, including those which are part of the formula which determine practice revenue under the NHS contracts. If we assume that we thereby controlled for any tendency for female GPs to be in less lucrative practices, then any remaining gender differences in income arise from the profit sharing rules within practices.[15]

GPs undertake a wide range of services which have different effects on the well being of their patients and on the income of the practice. The within practice profit sharing rules can lead to gender differences in pay in three ways. First, male GPs are paid more for any mix of activities. This is clearly gender discrimination. Second, female GPs could have lower incomes because they undertake a mix of activities which generates a lower share of the practice profit. Whether this is discrimination depends on whether female GPs choose their less profitable activity mix or whether it is assigned to them.

There is considerable evidence that male and female GPs have different consultation styles and produce different output mixes (Britt et al, 1996; Brink-Muinen et al, 2006; Langwell, 1982; Wilson, 1991; Roter et al, 1991; Roter et al, 2002). Rizzo and Zeckhauser (2007)

[15] There is a labour economics literature (Meng, 2004) which directly examines pay variations within firms and between firms. We have data on 1902 GPs in 1583 practices but only 144 practices have both male and female respondents which is too few to support a within practice analysis.

report differences in male and female physician attitudes to income generation. More generally, experiments suggest different attitudes of men and women to competition and cooperation in teams (Gneezy et al, 2003; Niederle and Vesterlund, 2005; Ivanova-Stenzel and Kubler, 2005; Flory et al, 2010).

We require a means of distinguishing between the three possible reasons (discrimination via differential rewards, discrimination by assignment of less financially rewarding activities, differences in preferences and productivity between male and female GPs) for gender related GP income differences.

*Mixed and single sex practice tests*
Within practice gender discrimination in profit sharing cannot arise 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 hours, experience, and the exogenous factors affecting total practice income, must be due to differences in preferences or productivity in income generation, not gender discrimination. 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.

To implement this test we estimate the log income model:

$$\ln y_i = \beta_0 + \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + F_i\left(\alpha_0 + \mathbf{x}_{1i}'\boldsymbol{\alpha}_1\right) + U_i\left[\gamma_0 + \mathbf{x}_{1i}'\boldsymbol{\gamma}_1 + F_i\left(\delta_0 + \mathbf{x}_{1i}'\boldsymbol{\delta}_1\right)\right] + \varepsilon_i \qquad (5)$$

where $F_i = 1$ if the GP is female, $U_i = 1$ if the practice GPs are all of the same gender, $\mathbf{x}_1$ 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 $\mathbf{x}_2$ is a vector of covariates whose effects are the same for all GPs.

We use (5) to estimate three gender income differences
(a) The conditional mean difference in log income between females in female only and mixed practices is

$$y_{ff} - y_{fm} \equiv \gamma_0 + \delta_0 + \overline{\mathbf{x}}_1^{F\prime}(\boldsymbol{\gamma}_1 + \boldsymbol{\delta}_1) \qquad (6)$$

where $\overline{\mathbf{x}}_1^F$ is the mean of $\mathbf{x}_1$ for 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 whatever the relative productivity or taste differences of male and female GPs. (And conversely if there is pro-female discrimination.) Thus pro-male discrimination implies that $y_{ff} - y_{fm} > 0$.

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

$$y_{mm} - y_{mf} \equiv \gamma_0 + \overline{\mathbf{x}}_1^{M\prime}\boldsymbol{\gamma}_1 \qquad (7)$$

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 $y_{mm} - y_{mf} < 0$.

(c) The conditional mean difference in log income between a female in an all female practice and a male in an all male practice, each with otherwise average female characteristics is

$$y_{ff} - y_{mm} \equiv \left[ \alpha_0 + \delta_0 + \overline{\mathbf{x}}_1^{F\prime}(\boldsymbol{\alpha}_1 + \boldsymbol{\delta}_1) \right] \qquad (8)$$

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

$$y_{fm} - y_{mf} \equiv \left[ \alpha_0 + \overline{\mathbf{x}}_1^{\prime F} \boldsymbol{\alpha}_1 \right] \qquad (9)$$

This difference may be due to taste or productivity differences as well as pro-male discrimination. Using (8) the extent of pro-male discrimination is measured by

$$(y_{fm} - y_{mf}) - (y_{ff} - y_{mm}) = -(\delta_0 + \overline{\mathbf{x}}_1^{\prime F} \boldsymbol{\delta}_1) \qquad (10)$$

which will be negative if there is pro-male discrimination.

The direct test based on (10) is vulnerable to two difficulties. There are relatively few female GPs in all female practices (35 of 753 female GPs) so that, although males in all male practices have larger incomes than females in all female practices, the difference ($y_{ff} - y_{mm}$) may be imprecisely estimated. The second difficulty is that if the practice characteristics in (5) do not control adequately for selection of female GPs into practices which are less profitable, then the difference ($y_{ff} - y_{mm}$) will overstate the difference in productivities. Such selection may be voluntarily, for example because practices offer different mixes of pay and working conditions which suit GPs with childcare commitments, or it may be discriminatory.

Our second and third direct tests of gender discrimination are not vulnerable to these difficulties.

*Test using the proportion of female GPs*

If there is gender discrimination against female GPs, it is plausible that there will be less of it in partnerships with a higher proportion of female GPs. Bell (2005) found that the pay of senior female executives was higher in companies with a larger share of female board members. We therefore estimate a model of the relationship between the gender mix in the practice and GP income as tests of gender discrimination:

$$\ln y_i = \beta_0 + \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + F_i\left(\alpha_0 + \mathbf{x}_{1i}'\boldsymbol{\alpha}_1\right) + \gamma_0 Femprop_i + \delta_0 Femprop_i F_i + \varepsilon_i \qquad (11)$$

where $Femprop_i$ is the proportion of female GPs in the practice to which GP $i$ belongs. If there is pro-male discrimination then $\delta_0 > 0$. Notice that even if there is discrimination in the selection of female GPs into less lucrative practices, so that practices with higher proportions of female GPs are potentially less profitable, the sign of $\delta_0$ is determined by the effect of gender mix on the sharing of a given profit between male and female GPs, not the effect of gender mix on practice profit.

*Test using the gender of senior partner*

Bell (2005) found that female executives in firms with a female chief executive or corporate chair have higher pay than comparable women in firms with male chief executives or corporate chairs. The general practice equivalent of the chair or chief executive is the senior partner. For our third set of direct discrimination we look at the effect of the gender of the senior partner on the incomes of the other GPs, male and female, in the practice. We estimate

$$\ln y_i = \beta_0 + \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + F_i\left(\alpha_0 + \mathbf{x}_{1i}'\boldsymbol{\alpha}_1\right) + Senior_i(\varphi_0 + \varphi_1 F_i)$$

$$+(1-Senior_i)(\delta_2 FemSenior_i + \delta_3 FemSenior_i F_i) + \varepsilon_i \qquad (12)$$

where $F_i$=1 if the GP is female, $Senior_i$ = 1 if GP $i$ is the senior partner, $FemSenior_i$=1 if the senior partner is female. If being in a practice in which the senior partner is female increases the income of female GPs relative to male GPs then $\delta_3 > 0$. As with the test for the effect of gender mix on the relative income of female GPs, even if practices with senior female GPs have lower incomes, $\delta_3$ is determined by the effect of a female senior partner on the sharing of a given profit between male and female GPs.

## 6.2 Indirect tests for discrimination

If female GPs face pay discrimination we would expect them to be less satisfied with their jobs than male GPs with similar incomes and hours. The Worklife Survey has a set of well validated questions about job satisfaction, both about overall job satisfaction "Taking everything into consideration, how do you feel about your job?" and about satisfaction with pay ("Please indicate how satisfied you are with each of the following aspects of your job" with one of the aspects being "Your remuneration".) Responses are on a Likert scale labeled from 1 (Extremely dissatisfied) to 7 (Extremely satisfied). We estimate ordered probit models for overall job satisfaction and for satisfaction with remuneration, with satisfaction assumed to depend on income, hours, experience, whether salaried, gender, ethnicity and country of qualification:

$$s_i = \beta_0 + \beta_1 y_i + \beta_2 h_i + \beta_3 years_i + \beta_4 salaried_i + \delta_1 F_i + \delta_2 Nonwhite_i + \delta_3 NonUKqual_i + \varepsilon_i \quad (13)$$

We assume that if female GPs are discriminated against then they are more likely to report lower levels of satisfaction than male GPs with the same contract, income, hours, ethnicity and country of qualification.

One objection to the comparison of satisfaction levels between male and female GPs as an indirect test of discrimination is that women in general report higher job satisfaction than men because their job expectations are lower (Clark, 1997). Clark (1997) argues that the difference in expectations between men and women will be smaller for better educated workers in professional jobs. He finds that in professional occupations there is no significant gender difference in job satisfaction or satisfaction with pay. Plausibly, therefore gender differences in job expectations will not be a source of higher job satisfaction for highly educated female GPs and discrimination will manifest itself as $\delta_1 < 0$.

The Worklife Survey also asks GPs what they consider to an adequate income for their job given their experience and career stage: "Considering your experience and career stage, what do you consider to be an adequate income from your job as a GP, after expenses but before taxes? £…,000 per year". This question is less vulnerable than the satisfaction questions to the possibility that female GPs have been conditioned to expect lower pay than men for similar work.

If female GPs suffer pay discrimination we expect them to feel that they are underpaid for the work they do and hence they are more likely than male GPs to report an adequate income which is greater than their actual income. We test for this in two ways. First, we run an OLS regression of the difference between adequate and actual income on experience, salaried status, gender, ethnicity and country of qualification

$$y_i^{adequate} - y_i^{actual} = \beta_0 + \beta_2 years_i + \beta_4 salaried_i + \delta_1 F_i + \delta_2 Nonwhite_i + \delta_3 NonUKqual_i + \varepsilon_i \quad (14)$$

Both $y_i^{adequate}$ and $y_i^{actual}$ are income band mid points, where we transform the continuous adequate income variable into the income bands used for actual income. Second, we measure the difference between adequate and actual income as an ordered categorical variable with eleven categories: adequate income four bands lower than actual income, three bands lower,…., six bands higher. We then estimate an ordered probit model using the same explanatories as in the OLS model (14).

# 7 Testing for within practice discrimination: results

## 7.1 Direct tests of within practice discrimination: results

*Unisex practice tests*

Table 4 reports results for the tests (6) to (8) from four versions of the regression model (5).[16] Model 1 uses all the explanatories in Table 1 with female interactions plus *U*, *U*F*, *U*hours*, and *U*hours*F*. Model 2 has the same explanatories as the reduced model 2 in Table 2 plus *U* and *U* interacted with all the female interaction variables in model 2, Table 2. Model 3 has the same explanatories as model 3 in Table 2 plus *U* and *U* interacted with female interactions. Model 4 is a reduced version of model 3 dropping variables which were or became insignificant at 10% but forcing in *U*, all the hours variables, and the non-white indicator.

(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 is positive and large for all four models and significant in three of them. This test suggests there may be pro-male discrimination in mixed practices since females in all female practices earn more than those in mixed gender practices.

(ii) All models show that male GPs in all male practices earn more than those in mixed practices ($y_{mm} - y_{mf} > 0$), and the difference is significant in three cases, suggesting that there is pro-female discrimination.

(iii) In all models the difference between the conditional mean log incomes of female and male GPs in mixed practices ($y_{fm} - y_{mf}$) is negative, large and in three of them is highly significant ($p < 0.001$). The conditional mean differences between the log incomes of female GPs in all female practices and male GPs in all male practices ($y_{ff} - y_{mm}$) is negative in three of the models and positive in one but is imprecisely estimated, with the lowest p value being 0.46.

The difference in income between female and male GPs, corrected for the estimated productivity or taste effect $(y_{fm} - y_{mf}) - (y_{ff} - y_{mm})$, is negative, varying between –17% (p=0.212) and –31.1% (p=0.113) but is significant only in model 4 (–28.8%, p=0.005).

The results from the direct tests for gender discrimination based on estimates of the effect of being in single sex or mixed sex practice provides some weak evidence for within practice discrimination against female GPs. However, the correction for tastes and/or productivity

[16] Full results are in Tables A4.1 and A4.2.

differences ($y_{ff} - y_{mm}$) is based on only small numbers of GPs in the relevant cells and is consequently poorly defined.

*Female GP proportion*

Table 5 reports the results for the models estimating the effects of the proportion of female GPs on GP income.[17] In model 1 in Table 5 GPs in practices with a higher proportion of female GPs have significantly lower incomes. This may be because female GPs select, voluntarily or otherwise, into practices with lower potential profit or it may reflect female GPs tastes for a less profitable mix of activities. The coefficient on the interaction of the female proportion with the female dummy is positive and very slightly larger than the negative coefficient on the female proportion. Thus female GPs in practices with a higher proportion of female GPs have higher income than male GPs. This suggests that gender mix affects the distribution of practice profit between male and female GPs and so is evidence for pro-male gender discrimination. However, the coefficient on the interaction term is not precisely estimated (p=0.259).

Models 2 and 5 in Table 5 replace the continuous female proportion variable with a more flexible categorical specification. Again a higher proportion of female GPs reduces the income of male GPs significantly. The negative effect is smaller for female GPs but again the difference between the effects on male and female GPs is not significant.

*Gender of senior partner*

Models 3 to 5 in Table 5 test for the effect of the gender of the senior partner on the incomes of male and female GPs. In all specifications, being the senior partner has no effect on the income of either male or female GPs because the models also include hours worked and senior partners work longer hours. The effect of having a female senior partner in the practice on the income of other GPs is insignificant for both male and female GPs.

## 7.2 Indirect tests for gender discrimination: results

Table 6 reports results from the indirect tests for gender discrimination based on job satisfaction, satisfaction with pay, and the difference between adequate and actual income. Female GPs are not more likely to report lower job satisfaction than male GPs (Model 1). They are however more likely to report greater satisfaction with their pay than male GP (Model 2). Interestingly, salaried GPs (whose incomes according to the models in Table 2 are around half those of other GPs for the same hours worked) are more likely to be dissatisfied with pay, as are non-white GPs whose pay is 6% less than otherwise similar GPs.

Models 3 and 4 show that whilst all types of GPs report that an adequate income for their job exceeds their actual income, the difference is not significantly greater for female GPs than for male GPs. Non-white GPs and overseas qualified GPs report significantly larger differences between adequate and actual income than white and UK qualified GPs.

[17] Full results are in Tables A5.1 and A5.2

# 8 Conclusions

Conditional on observed factors, female GPs have lower incomes, lower average wages (income/hours) but higher marginal rewards for hours. Whilst female income is nearly proportional to hours worked (elasticity = 0.91), the elasticity of male income with respect to hours is much smaller at around 0.29. This implies that models of male GP pay based on wages calculated as income/hours are misspecified since the calculated wage declines with hours worked. It also implies that, since female GPs work shorter hours (34.5 per week against male hours of 45.9), the difference in log wages (11%) is markedly less than the difference in log incomes (45%).

The unexplained difference in log incomes is around 30% and is insensitive to the specification of the log income regression. However, this unexplained difference is the sum of two terms which are very large in absolute value but of different signs: the differences in rewards to observable characteristics and differences in the regression constants. If there are omitted variables the difference in regression constants will include both differences in rewards to unobserved factors and differences in the means of the unobserved factors. The conventional decomposition which sums the differences in constants and differences in returns is thus a misleading measure of discrimination since it is impossible to determine whether the main component arises from differences in returns from differences in means.

We have proposed new tests for within practice income discrimination based on the comparison of GPs in practices with differing gender mixes. Tests based on comparisons of female and male incomes in single gender and mixed gender practice using our preferred model specification provide some evidence of pro-male discrimination but may be vulnerable to the small number of female GPs in all female practices.

If there is discrimination against female GPs we would expect it would less severe in practices with a female senior partner or with a higher proportion of female GPs. We find no effect of the gender of the senior GP on the incomes of male or female GPs. All GPs have smaller incomes in practices where the proportion of female GPs is higher but there is no significant differences in the effect on male and female GPs.

We also find that female GPs, compared to similar male GPs are more satisfied with their job overall, have the same satisfaction with income, and are not more likely to report that an adequate income for their job exceeds their actual income.

The results from our tests for discrimination provide only weak support for the existence of within practice discrimination. Our finding that GPs in practices with a higher proportion of female GPs have lower incomes could be due to selection of female GPs into less profitable practices but could also be due to gender differences in preferences or productivity across activity mixes. If there is gender discrimination it is surprising that it does not manifest itself in dissatisfaction with pay or in a larger difference between reported adequate and actual income for female GPs.

**Acknowledgements**. 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 2008 General Practitioner Worklife Survey was undertaken by Bonnie Sibbald, Mark Hann, Matt Sutton and Hugh Gravelle. We are grateful to Mark Dusheiko for providing additional data and for comments from two referees, from Steve Morris, Karen Mumford, Jo Swaffield, and from participants in seminars at Bergen, Sheffield, York, HEDG, the Melbourne Institute, the 3rd Portuguese Health Economics Association workshop.

## References

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.

Baker, R. 1996. Characteristics of practices, general practitioners and patients related to levels of patients' satisfaction with consultations. *British Journal of General Practice*. 46, 601–5.

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.

Bell, L. 2005. Women-led firms and the gender gap in top executive jobs. Institute for the Study of Labor. IZA Discussion Paper No. 1689.

Bertrand, M., Goldin, C., and Katz, L. 2009. Dynamics of the gender gap for young professionals in the corporate and financial sectors. NBER Working Paper 14861.

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.

Brink-Muinen, A., Dulmen, S., and Bensing, J. 2006. Have gender preferences and communication patterns changed? In Westert, G., Jabaaij, L., and Schellevis, F. (eds.). *Morbidity, Performance and Quality in Primary Care*, Radcliffe Publishing, Oxford.

Britt, H., Bhasale, A., Miles, D., Meza, A., Sayer, G. and Angelis, M. 1996. The sex of the general practitioner. A comparison of characteristics, patients, and medical conditions managed. *Medical Care*. 34, 403-415.

Clark, A. 1997. Job satisfaction and gender. Why are women so happy at work? *Labour Economics*, 4, 341-372.

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.

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.

Flory, J., Leibbrandt, A. and List, J. 2010. Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job entry decisions. *National Bureau of Economic Research Working Paper* No. 16546, November.

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.

Gneezy, U., Niederle, M. and Rustichini, A. 2003. Performance in competitive environments: gender differences. *Quarterly Journal of Economics*. 118, 1049-1074.

Gravelle, H. and Hole, M. 2008. Measuring and testing for gender discrimination in professions: the case of English GPs. *Department of Economics and Related Studies, Discussion Paper* 2008/28.

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

Hann, M., Reeves, D. and Roland, M. Patient Experience with Primary Care. An analysis of three waves of the IPSOS-MORI patient survey data. National Primary Care Research and Development Centre. June 2008

Hann, M., Goodie, R., Sutton, M., Gravelle, H. and Sibbald, B. 2009. Fifth National GP Worklife Survey. Final Report. http://www.npcrdc.ac.uk/PublicationResults.cfm?Type=2

Information Centre. 2009a. Prescriptions Dispensed in the Community. Statistics for 1998 to 2008: England. http://www.ic.nhs.uk/webfiles/publications/presdisp98-08/Prescriptions_Dispensed_in_the_Community_1998_2008_England_2.pdf

Information Centre. 2009b. General and Personal Medical Services England 1998-2008. http://www.ic.nhs.uk/statistics-and-data-collections/workforce/nhs-staff-numbers/nhs-staff-1998--2008-general-practice.

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.

Jenkins-Clarke, S., Carr-Hill, R., Dixon, P., and Pringle, M. Skill mix in primary care. University of York, *Centre for Health Economics Occasional Paper* 28, 1997.

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

Jones, F.L. 1983. On decomposing the wage gap: a critical comment on Blinder's method. *Journal of Human Resources*. 18, 126-130.

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.

Meng, X. 2004. Gender earnings gap: the role of firm specific effects. *Labour Economics*, 11, 555-573.

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.

Morris S, Goudie R, Sutton M, Gravelle H, Elliott R, Hole A, Ma A, Sibbald B, and Skatun D. 2010. Determinants of general practitioners' wages in England. *Health Economics*. 2010, in press.

Murphy-Cullen, C.L. and Larsen, L.C. 1984. Interaction between the socio-demographic variables of physicians and their patients: Its impact upon patient satisfaction. *Social Science & Medicine*. 19, 163-166.

National Audit Office. 2008. NHS Pay modernisation: New Contracts for General Practice Services in England. National Audit Office press office, London.

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

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

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.

Phillips, D. and Brooks, F. 1998. Women patients' preferences for female or male GPs. *Family Practice*. 15, 543-547.

Pope, G. and Burge, R. 1996. Economies of scale in physician practice *Medical Care Research and Review*, 53(4) 417-40.

Reyes, J. 2006. Do female physicians capture their scarcity value? The case of OB/GYNs. *National Bureau of Economic Research*, Working Paper 12528.

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.

Roland, M. 2004. Linking physicians pay to the quality of care – a major experiment in the United Kingdom. *New England Journal of Medicine*. 351, 1448–54.

Roter, D.L., Hall, J.A. and Aoki, Y. 2002. Physician gender effects in medical communication: a meta-analytic review. *JAMA*. 288, 756–764.

Roter, D., Lipkin, M. and Korsgaard, A. 1991. Sex differences in patient's and physicians' communication during primary care medical visits. *Medical care*, 29(11), 1083-1093.

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

Shih, Y.-C., T. and Konrad, T.R. 2007. Factors associated with the income distribution of full time physicians: a quantile regression approach. *Health Services Research*. 42, 5, 1895-1925.

Theurl, E., and Winner, H. 2010. The male-female gap in physician earnings: evidence from a public health insurance system. *Health Economics*. Published online DOI:10.1002/hec.1663.

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 Labor Economics*, 11, 417-441.

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

**Table 1. Descriptive Statistics**

| Name | Variable Definition | Female GPs Mean | Female GPs SD | Male GPs Mean | Male GPs SD | |
|---|---|---|---|---|---|---|
| Income | Annual earning in £000s | 77.03 | 32.61 | 110.69 | 30.46 | ** |
| Adequate inc | Adequate income for job in £000s | 99.47 | 35.48 | 131.09 | 30.52 | ** |
| Hours | Annual hours worked | 1580.56 | 520.92 | 2106.58 | 514.46 | ** |
| Hours pw | Weekly hours worked | 34.53 | 11.35 | 45.88 | 10.95 | ** |
| Wage | Wage (Income/Hours) | 51.65 | 35.63 | 58.33 | 58.15 | ** |
| Satisfaction | Overall job satisfaction (1 to 7) | 4.77 | 1.25 | 4.69 | 1.36 | |
| Pay satisfaction | Satisfaction with pay (1 to 7) | 4.83 | 1.52 | 4.78 | 1.45 | |
| Service | Decades of reckonable service | 2.04 | 0.74 | 2.31 | 0.80 | ** |
| Age | Age of GP | 46.00 | 7.37 | 49.37 | 7.49 | ** |
| NonWhite | Non-white | 0.10 | | 0.14 | | ** |
| NonUK | Non-UK qualified | 0.08 | | 0.12 | | ** |
| Child<4 | Youngest child under 4 | 0.19 | | 0.11 | | ** |
| Child4-18 | Youngest child between 4 and 18 | 0.49 | | 0.49 | | |
| SpouseFT | Spouse/partner works full time | 0.64 | | 0.18 | | ** |
| Traveltime | Travel time to practice | 18.33 | 11.61 | 15.46 | 10.79 | ** |
| Senior | Senior GP | 0.19 | | 0.39 | | ** |
| Salaried | Salaried GP | 0.14 | | 0.04 | | ** |
| LISI | Low Income Scheme Index (LISI) score | 9.57 | 6.10 | 9.54 | 5.95 | |
| Noqual | Proportion of adults with no qualification | 0.28 | 0.08 | 0.29 | 0.08 | ** |
| List | Practice size in '000s | 9.31 | 4.20 | 9.59 | 4.65 | |
| ListperGP | Practice size per GP in 000's | 1.68 | 0.40 | 1.71 | 0.45 | |
| AgeSex | Age and Sex capitation adjustment | 2.54 | 0.21 | 2.56 | 0.21 | * |
| Nursing home | Proportion of population in nursing home | 0.01 | 0.01 | 0.01 | 0.01 | |
| Needs | Additional needs capitation index | 96.33 | 6.72 | 96.63 | 6.62 | |
| QOF points | Number of QOF points 2007/8 | 987.13 | 22.71 | 983.29 | 29.33 | ** |
| MFFstaff | Index of cost of staff | 1.00 | 0.10 | 0.99 | 0.09 | ** |
| MFFpremises | Index of cost of premises | 0.98 | 0.75 | 0.86 | 0.61 | ** |
| Dispensing | Dispensing practice | 0.17 | | 0.18 | | |
| PMS | PMS practice | 0.43 | | 0.46 | | |
| Unisex | GP of same gender as all GP providers | 0.05 | | 0.17 | | ** |
| Femprop | Proportion of female GPs in practice | 0.49 | 0.19 | 0.32 | 0.20 | ** |
| FemSenior | Senior GP in practice is female | 0.36 | | 0.13 | | ** |
| NonSenior x FemSenior | Non-senior GP in practice with female senior GP | 0.17 | | 0.13 | | * |
| Amenities | Good local amenities | 0.71 | | 0.68 | | |
| Rurality | Rurality | 0.02 | | 0.02 | | |
| Housing | High local housing costs | 0.70 | | 0.66 | | * |
| GOODschools | Good local schools | 0.74 | | 0.70 | | |
| Optout | Practice opted out of out-of-hours | 1.14 | | 1.11 | | * |
| Extendedhours | Practice offers extended opening hours | 1.55 | | 1.58 | | |
| Competition | Number of practices within 1km | 2.15 | 2.31 | 2.15 | 2.30 | |
| GOR1 | North East | 0.07 | 0.26 | 0.07 | 0.26 | ** |
| GOR2 | North West | 0.13 | 0.34 | 0.11 | 0.32 | ** |
| GOR3 | Yorkshire | 0.11 | 0.32 | 0.10 | 0.30 | ** |
| GOR4 | East Midlands | 0.08 | 0.27 | 0.09 | 0.28 | ** |

| | | | | | | |
|---|---|---|---|---|---|---|
| GOR5 | West Midlands | 0.07 | 0.26 | 0.11 | 0.31 | ** |
| GOR6 | East | 0.09 | 0.29 | 0.12 | 0.33 | ** |
| GOR7 | London | 0.14 | 0.35 | 0.09 | 0.29 | ** |
| GOR8 | South East | 0.19 | 0.39 | 0.17 | 0.38 | ** |
| GOR9 | South West | 0.11 | 0.32 | 0.14 | 0.34 | ** |
| | N | 734 | | 1168 | | |

Significant levels of differences in means or proportions: *p<0.05, ** p<0.01; t tests except Chi square for GOR

**Table 2. Log wage and log income regressions**

| | Wage model | | Income models | | | |
|---|---|---|---|---|---|---|
| | Model 1 | | Model 2 | | Model 3 | |
| | | | Full interactions | | Restricted interactions | |
| | Coef. | p-value | Coef. | p-value | Coef. | p-value |
| Hours x Female | | | 0.906*** | (0.000) | 0.912*** | (0.000) |
| Hours x Male | | | 0.286*** | (0.000) | 0.282*** | (0.000) |
| Female (F) | -1.593 | (0.802) | -6.530 | (0.290) | -6.030** | (0.003) |
| Service (10-20yrs) | 0.100 | (0.093) | 0.067 | (0.136) | 0.049 | (0.294) |
| F x Service (10-20yrs) | -0.152 | (0.074) | -0.165 | (0.050) | -0.138 | (0.084) |
| Service (20-30yrs) | 0.103 | (0.093) | 0.072 | (0.092) | 0.054 | (0.213) |
| F x Service (20-30yrs) | -0.180* | (0.038) | -0.206* | (0.014) | -0.174* | (0.026) |
| Service (>30yrs) | 0.147* | (0.017) | 0.063 | (0.151) | 0.043 | (0.327) |
| F x Service (>30yrs) | -0.150 | (0.098) | -0.105 | (0.227) | -0.065 | (0.438) |
| NonWhite | -0.045 | (0.381) | -0.060 | (0.079) | -0.055 | (0.110) |
| F x NonWhite | 0.085 | (0.287) | 0.126 | (0.078) | 0.115 | (0.100) |
| Salaried | -0.192** | (0.002) | -0.426*** | (0.000) | -0.495*** | (0.000) |
| F x Salaried | -0.112 | (0.172) | -0.093 | (0.372) | | |
| ln List | 0.197 | (0.158) | 0.105 | (0.341) | 0.081 | (0.463) |
| F x ln List | -0.494 | (0.055) | -0.455 | (0.074) | -0.443 | (0.064) |
| List 2 | -0.033 | (0.324) | -0.013 | (0.616) | -0.008 | (0.769) |
| F x List 2 | 0.118* | (0.043) | 0.115* | (0.046) | 0.110* | (0.047) |
| ln ListperGP | -1.750* | (0.015) | -1.948*** | (0.000) | -1.962*** | (0.001) |
| F x ln ListperGP | 0.205 | (0.901) | -0.026 | (0.986) | | |
| ListperGP 2 | 0.126* | (0.012) | 0.140*** | (0.000) | 0.140*** | (0.000) |
| F x ListperGP 2 | -0.022 | (0.845) | -0.001 | (0.991) | | |
| ln AgeSex | -0.125 | (0.492) | -0.008 | (0.950) | -0.022 | (0.866) |
| F x ln AgeSex | -0.269 | (0.361) | -0.515 | (0.065) | -0.495* | (0.046) |
| ln Needs | -0.352 | (0.078) | -0.445** | (0.010) | -0.382* | (0.027) |
| F x ln Needs | 0.448 | (0.216) | 0.628 | (0.098) | 0.459 | (0.205) |
| Dispensing | 0.159*** | (0.000) | 0.156*** | (0.000) | 0.175*** | (0.000) |
| F x Dispensing | 0.016 | (0.773) | 0.060 | (0.271) | | |
| PMS | 0.107*** | (0.000) | 0.115*** | (0.000) | 0.106*** | (0.000) |
| F x PMS | -0.042 | (0.281) | -0.030 | (0.431) | | |
| Optout | 0.115* | (0.021) | 0.091** | (0.002) | 0.104*** | (0.000) |
| F x Optout | -0.013 | (0.856) | 0.026 | (0.653) | | |
| Amenities | 0.061** | (0.009) | 0.023 | (0.165) | 0.043* | (0.016) |
| F x Amenities | 0.013 | (0.744) | 0.057 | (0.160) | | |
| Rurality | 0.150* | (0.028) | 0.116* | (0.029) | 0.049 | (0.306) |
| F x Rurality | -0.238* | (0.025) | -0.148 | (0.121) | | |
| Traveltime | -0.001 | (0.470) | -0.001 | (0.068) | -0.001 | (0.066) |
| F x Traveltime | -0.000 | (0.953) | -0.000 | (0.918) | | |
| Femprop | -0.122 | (0.062) | -0.116* | (0.020) | -0.070 | (0.159) |
| F x Femprop | 0.101 | (0.407) | 0.123 | (0.305) | | |
| GOR6 | -0.021 | (0.606) | -0.031 | (0.231) | -0.028 | (0.286) |
| F x GOR6 | 0.064 | (0.395) | 0.035 | (0.655) | 0.027 | (0.729) |
| GOR7 | -0.096 | (0.101) | -0.042 | (0.431) | -0.040 | (0.454) |

| | | | | | | |
|---|---|---|---|---|---|---|
| F x GOR7 | 0.214* | (0.010) | 0.177* | (0.033) | 0.171* | (0.028) |
| GOR8 | -0.051 | (0.120) | -0.039 | (0.290) | -0.036 | (0.327) |
| F x GOR8 | 0.041 | (0.512) | 0.008 | (0.910) | 0.000 | (0.998) |
| GOR9 | -0.093** | (0.003) | -0.112*** | (0.000) | -0.111*** | (0.000) |
| F x GOR9 | 0.085 | (0.094) | 0.105* | (0.049) | 0.093 | (0.079) |
| Constant | 11.176*** | (0.000) | 10.963*** | (0.000) | 10.821*** | (0.000) |
| N | 1902 | | 1902 | | 1902 | |
| BIC | 1936.478 | | 1777.693 | | 1714.596 | |
| F-test | | | 18.85 *** | | 21.07*** | |
| Endogeneity(DWH) | | | 4.748** | | 4.452** | |
| Hansen J stats p-value | | | 0.140 | | 0.158 | |
| Adjusted $R^2$ | 0.119 | | 0.506 | | 0.505 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001

**Table 3A - Decomposition of differences in log pay: male coefficients as counterfactual**

| | | Wage model | | Income models | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | OLS (Table 2, Model 1) | | 2SLS, full set of female interactions (Table 2, Model 2) | | 2SLS, parsimonious set of female interactions (Table 2, Model 3) | | OLS, full set of female interactions | |
| | | Diff | t stat | Diff | t stat | Diff | t stat | Diff | t stat |
| Diff in constants | $\hat{\beta}_0^M - \hat{\beta}_0^F$ | 1.78 | 0.28 | 6.59 | 1.07 | 6.01 | 2.95 | 4.00 | 0.77 |
| Diff in returns | $\overline{\mathbf{x}}^{F\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)$ | -1.72 | -0.27 | -6.29 | -1.02 | -5.71 | -2.81 | -3.70 | -0.71 |
| Unexplained difference | $\hat{\beta}_0^M - \hat{\beta}_0^F + \overline{\mathbf{x}}^{F\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)$ | 0.06 | 3.03 | 0.30 | 9.73 | 0.30 | 9.38 | 0.30 | 9.80 |
| Difference in characteristics (Explained difference) | $(\overline{\mathbf{x}}^{M\prime} - \overline{\mathbf{x}}^{F\prime})\hat{\boldsymbol{\beta}}^M$ | 0.05 | 3.88 | 0.16 | 7.15 | 0.15 | 6.54 | 0.16 | 7.15 |
| Total difference | $\overline{\ln z^M} - \overline{\ln z^F}$ | 0.11 | 6.51 | 0.45 | 25.93 | 0.45 | 25.80 | 0.45 | 26.47 |

**Table 3B. Decompositions of differences in log pay: female coefficients as counterfactual**

| | | Wage model | | Income models | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | OLS (Table 2, Model 1) | | 2SLS, full set of female interactions (Table 2, Model 2) | | 2SLS, parsimonious set of female interactions (Table 2, Model 3) | | OLS, full set of female interactions | |
| | | Diff | t stat | Diff | t stat | Diff | t stat | Diff | t stat |
| Diff in constants | $\hat{\beta}_0^M - \hat{\beta}_0^F$ | 1.78 | 0.28 | 6.59 | 1.07 | 6.01 | 2.95 | 4.00 | 0.77 |
| Diff in returns | $\overline{\mathbf{x}}^{M\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)$ | -1.71 | -0.27 | -6.47 | -1.05 | -5.91 | -2.87 | -3.81 | -0.73 |
| Unexplained difference | $\hat{\beta}_0^M - \hat{\beta}_0^F + \overline{\mathbf{x}}^{M\prime}(\hat{\boldsymbol{\beta}}^M - \hat{\boldsymbol{\beta}}^F)$ | 0.07 | 2.64 | 0.11 | 2.57 | 0.10 | 2.36 | 0.19 | 5.80 |
| Difference in characteristics (Explained difference) | $(\overline{\mathbf{x}}^{M\prime} - \overline{\mathbf{x}}^{F\prime})\hat{\boldsymbol{\beta}}^F$ | 0.04 | 2.16 | 0.34 | 8.32 | 0.36 | 9.07 | 0.26 | 7.97 |
| Total difference | $\overline{\ln z^M} - \overline{\ln z^F}$ | 0.11 | 6.51 | 0.45 | 25.93 | 0.45 | 25.80 | 0.45 | 26.47 |

**Table 4 . Direct tests for differences in mean income by gender mix of practice**

| Difference in conditional means | Model 1 Coef. | Model 1 p | Model 2 Coef. | Model 2 p | Model 3 Coef. | Model 3 p | Model 4 Coef. | Model 4 p |
|---|---|---|---|---|---|---|---|---|
| $y_{ff} - y_{fm}$ | 0.219 | 0.035 | 0.411 | 0.029 | 0.221 | 0.354 | 0.397 | 0.000 |
| $y_{mm} - y_{mf}$ | 0.055 | 0.064 | 0.101 | 0.000 | 0.105 | 0.000 | 0.086 | 0.004 |
| $y_{ff} - y_{mm}$ | -1.806 | 0.771 | 0.014 | 0.944 | -0.181 | 0.456 | -0.011 | 0.909 |
| $y_{fm} - y_{mf}$ | -1.976 | 0.750 | -0.298 | 0.000 | -0.297 | 0.000 | -0.299 | 0.000 |
| $(y_{fm} - y_{mf}) - (y_{ff} - y_{mm})$ | -0.170 | 0.212 | -0.311 | 0.113 | -0.116 | 0.635 | -0.288 | 0.005 |
| N | 1626 | | 1902 | | 1902 | | 1902 | |
| BIC | 1819.22 | | 2024.99 | | 1878.14 | | 1675.19 | |

$y_{ff}$ ($y_{mm}$) conditional mean income of female (male) GP in single sex practice; $y_{fm}$ ($y_{mf}$) conditional mean income of female (male) GP in mixed sex practice. Full results of 2SLS income models are in Appendix.
Model 1 has all the variables from Table 1 (except adequate income, hours per week, wage, job satisfaction, satisfaction with pay, age of GP, proportion of female GPs and gender of the senior GP in practice ) with a full set of female interactions, plus unisex practice dummy variable, female GPs in unisex practices, female GPs annual hours in unisex practices and male GPs annual hours in unisex practices.
Model 2 has all the variables from income model 2, Table 2, plus unisex practice dummy variable, female GPs in unisex practice, and full set of unisex practice interactions.
Model 3 has all the variables from income model 3, Table 2, plus unisex practice dummy variable, female GPs in unisex practice and a set of unisex practice interactions with all the variables with a female interaction in model 3 (Table 2).
Model 4 is a nested version of model 3.

**Table 5. Direct tests of discrimination: effects of being senior GP, gender of senior GP, and proportion of female GPs on log income**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Coef | p | Coef | p | Coef | p | Coef | p | Coef | |
| Senior | | | | | 0.037 | 0.155 | 0.040 | 0.124 | 0.040 | 0.123 |
| F*Senior | | | | | -0.004 | 0.936 | -0.006 | 0.905 | -0.001 | 0.982 |
| | | | | | | | | | | |
| Prop female GP | -0.117 | 0.023 | | | | | -0.138 | 0.010 | | |
| F*Prop female GP | 0.137 | 0.259 | | | | | 0.136 | 0.283 | | |
| | | | | | | | | | | |
| Prop female GP ∈ (0.25-0.50] | | | -0.062 | 0.001 | | | | | -0.065 | 0.001 |
| Prop female GP ∈ (0.50-0.75] | | | -0.073 | 0.057 | | | | | -0.089 | 0.028 |
| Prop female GP > 0.75 | | | -0.144 | 0.052 | | | | | -0.154 | 0.039 |
| F*Prop female GP ∈ (0.25-0.50] | | | -0.019 | 0.774 | | | | | -0.013 | 0.847 |
| F*Prop female GP ∈ (0.50-0.75] | | | 0.071 | 0.361 | | | | | 0.094 | 0.247 |
| F*Prop female GP > 0.75 | | | 0.052 | 0.679 | | | | | 0.035 | 0.779 |
| | | | | | | | | | | |
| Female senior GP*non senior | | | | | 0.002 | 0.926 | 0.024 | 0.380 | 0.027 | 0.329 |
| F*Female senior GP*non senior | | | | | -0.010 | 0.842 | -0.032 | 0.556 | -0.038 | 0.472 |
| | | | | | | | | | | |
| BIC | 1743.9 | | 1758.7 | | 1744.1 | | 1752.9 | | 1766.533 | |
| N | 1583 | | 1583 | | 1576 | | 1576 | | 1576 | |

Results from 2SLS models with the same controls as Model 3, Table 2: Full results in appendix.

**Table 6. Indirect tests of discrimination: job satisfaction; pay satisfaction; differences between adequate and actual income.**

| | Overall satisfaction | | Satisfaction with pay | | Diff. adequate and actual income | | | |
|---|---|---|---|---|---|---|---|---|
| | Ordered Probit | | | | OLS | | Ordered Probit | |
| ln Income | 0.363*** | (0.000) | 0.640*** | (0.000) | | | | |
| ln Hours | -0.562*** | (0.000) | -0.538*** | (0.000) | | | | |
| Service (10-20yrs) | -0.163 | (0.102) | 0.120 | (0.208) | 2.889 | (0.319) | 0.127 | (0.262) |
| Service (20-30yrs) | -0.207* | (0.040) | 0.134 | (0.157) | 3.976 | (0.169) | 0.162 | (0.153) |
| Service (>30yrs) | 0.0159 | (0.887) | 0.299** | (0.005) | 2.652 | (0.391) | 0.103 | (0.392) |
| Salaried | -0.102 | (0.359) | -0.241* | (0.029) | 2.681 | (0.292) | 0.107 | (0.281) |
| Female | 0.0724 | (0.201) | 0.201*** | (0.001) | 1.784 | (0.190) | 0.0587 | (0.272) |
| NonWhite | -0.108 | (0.189) | -0.195* | (0.017) | 12.36*** | (0.000) | 0.497*** | (0.000) |
| NonUK | -0.0781 | (0.382) | -0.157 | (0.054) | 8.387*** | (0.000) | 0.338*** | (0.000) |
| Constant | | | | | 14.82*** | (0.000) | | |
| cut1 | -4.861*** | (0.000) | -2.969*** | (0.000) | | | -2.673*** | (0.000) |
| cut2 | -4.235*** | (0.000) | -2.435*** | (0.000) | | | -2.212*** | (0.000) |
| cut3 | -3.672*** | (0.000) | -1.924** | (0.005) | | | -1.929*** | (0.000) |
| cut4 | -3.084*** | (0.000) | -1.359* | (0.046) | | | -1.369*** | (0.000) |
| cut5 | -2.184** | (0.001) | -0.660 | (0.333) | | | -0.0334 | (0.768) |
| cut6 | -1.061 | (0.120) | 0.388 | (0.569) | | | 0.984*** | (0.000) |
| cut7 | | | | | | | 1.771*** | (0.000) |
| cut8 | | | | | | | 2.513*** | (0.000) |
| cut9 | | | | | | | 3.309*** | (0.000) |
| cut10 | | | | | | | 3.647*** | (0.000) |
| N | 1955 | | 1957 | | 1791 | | 1791 | |
| $R^2$ | | | | | 0.044 | | | |
| $\rho^2$ | 0.015 | | 0.03 | | | | 0.016 | |

p-values in parentheses, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

## Appendix

### Table A1. Full model with full set of female interactions

| (first and second stage) | **Log income** | | **Log hours** | |
|---|---|---|---|---|
| Hours x Female | 0.932*** | (0.000) | | |
| Female | -17.405 | (0.228) | -5.624 | (0.596) |
| Hours x Male | 0.301*** | (0.001) | -0.000 | (0.487) |
| Service (10-20yrs) | 0.095* | (0.044) | 0.000 | (0.481) |
| F x Service (10-20yrs) | -0.147 | (0.114) | -0.078 | (0.189) |
| Service (20-30yrs) | 0.095* | (0.030) | 0.000 | (0.675) |
| F x Service (20-30yrs) | -0.184 | (0.053) | -0.097 | (0.151) |
| Service (>30yrs) | 0.079 | (0.093) | 0.000 | (0.635) |
| F x Service (>30yrs) | -0.050 | (0.613) | -0.164* | (0.026) |
| NonWhite | -0.057 | (0.152) | -0.000 | (0.561) |
| F x NonWhite | 0.150 | (0.080) | -0.046 | (0.513) |
| NonUK | -0.003 | (0.939) | 0.000 | (0.518) |
| F x NonUK | -0.092 | (0.266) | 0.138** | (0.010) |
| Senior | 0.005 | (0.852) | -0.000 | (0.609) |
| F x Senior | 0.010 | (0.878) | 0.163*** | (0.000) |
| Salaried | -0.387*** | (0.000) | 0.000 | (0.663) |
| F x Salaried | -0.120 | (0.355) | -0.421*** | (0.000) |
| ln LISI | -0.019 | (0.524) | -0.000 | (0.723) |
| F x ln LISI | 0.022 | (0.742) | -0.042 | (0.307) |
| ln Noqual | 0.005 | (0.937) | 0.001 | (0.460) |
| F x ln Noqual | -0.056 | (0.611) | 0.154* | (0.040) |
| ln List | 0.094 | (0.426) | 0.001 | (0.666) |
| F x ln List | -0.510 | (0.090) | -0.012 | (0.955) |
| List 2 | -0.007 | (0.809) | -0.000 | (0.666) |
| F x List 2 | 0.124 | (0.065) | 0.030 | (0.523) |
| ln ListperGP | -2.471** | (0.003) | -1.309 | (0.453) |
| F x ln ListperGP | 2.662 | (0.450) | 3.366 | (0.210) |
| ListperGP 2 | 0.174** | (0.002) | 0.178 | (0.452) |
| F x ListperGP 2 | -0.183 | (0.443) | -0.218 | (0.230) |
| ln AgeSex | -0.215 | (0.225) | 0.002 | (0.481) |
| F x ln AgeSex | -0.421 | (0.293) | -0.044 | (0.870) |
| ln Nursing home | 0.011 | (0.365) | -0.000 | (0.500) |
| F x ln Nursing home | -0.038 | (0.219) | 0.023 | (0.337) |
| ln Needs | -0.340 | (0.242) | -0.002 | (0.532) |
| F x ln Needs | 0.770 | (0.196) | -0.553 | (0.170) |
| ln QOF points | 0.186 | (0.742) | 0.003 | (0.511) |
| F x ln QOF points | -0.039 | (0.963) | 0.455 | (0.399) |
| Dispensing | 0.157*** | (0.000) | 0.000 | (0.490) |
| F x Dispensing | 0.060 | (0.353) | 0.042 | (0.339) |
| PMS | 0.125*** | (0.000) | 0.000 | (0.463) |
| F x PMS | -0.002 | (0.968) | -0.016 | (0.594) |
| Optout | 0.106*** | (0.000) | 0.000 | (0.609) |
| F x Optout | -0.003 | (0.956) | -0.008 | (0.857) |
| Extendedhours | 0.011 | (0.341) | -0.000 | (0.598) |
| F x Extendedhours | 0.001 | (0.963) | 0.007 | (0.702) |

| | | | | |
|---|---|---|---|---|
| Housing | 0.006 | (0.773) | -0.000 | (0.921) |
| F x Housing | -0.017 | (0.726) | 0.019 | (0.563) |
| GOODschools | 0.017 | (0.466) | -0.000 | (0.457) |
| F x GOODschools | 0.010 | (0.846) | -0.028 | (0.441) |
| Amenities | 0.013 | (0.520) | 0.000 | (0.426) |
| F x Amenities | 0.082 | (0.137) | -0.045 | (0.226) |
| Rurality | 0.135* | (0.022) | 0.000 | (0.467) |
| F x Rurality | -0.149 | (0.154) | 0.084 | (0.295) |
| Traveltime | -0.001 | (0.082) | -0.000 | (0.536) |
| F x Traveltime | -0.001 | (0.756) | 0.000 | (0.918) |
| Competition | -0.002 | (0.690) | 0.000 | (0.518) |
| F x Competition | 0.001 | (0.929) | 0.018* | (0.027) |
| ln MFFstaff | 0.307 | (0.358) | -0.001 | (0.645) |
| F x ln MFFstaff | -0.537 | (0.425) | 0.671 | (0.079) |
| ln MFFpremises | -0.048 | (0.272) | 0.000 | (0.573) |
| F x ln MFFpremises | -0.047 | (0.590) | 0.029 | (0.590) |
| Femprop | -0.100 | (0.053) | 0.000 | (0.558) |
| F x Femprop | 0.070 | (0.604) | 0.156 | (0.074) |
| GOR2 | 0.038 | (0.368) | -0.000 | (0.727) |
| F x GOR2 | 0.131 | (0.254) | -0.127 | (0.051) |
| GOR3 | 0.035 | (0.491) | 0.000 | (0.868) |
| F x GOR3 | 0.141 | (0.221) | -0.093 | (0.165) |
| GOR4 | 0.027 | (0.596) | -0.001 | (0.464) |
| F x GOR4 | 0.148 | (0.229) | -0.085 | (0.300) |
| GOR5 | 0.058 | (0.207) | -0.000 | (0.684) |
| F x GOR5 | 0.144 | (0.241) | -0.084 | (0.256) |
| GOR6 | 0.023 | (0.654) | -0.000 | (0.495) |
| F x GOR6 | 0.238 | (0.146) | -0.266** | (0.002) |
| GOR7 | -0.001 | (0.988) | -0.000 | (0.645) |
| F x GOR7 | 0.464* | (0.020) | -0.234* | (0.034) |
| GOR8 | 0.005 | (0.935) | -0.000 | (0.670) |
| F x GOR8 | 0.208 | (0.199) | -0.228* | (0.013) |
| GOR9 | -0.057 | (0.230) | 0.000 | (0.565) |
| F x GOR9 | 0.245* | (0.040) | -0.109 | (0.105) |
| ListperGP 3 | | | -0.008 | (0.451) |
| F x Child<4 | | | -0.085 | (0.141) |
| F x Child4-18 | | | -0.172*** | (0.000) |
| F x SpouseFT | | | -0.034 | (0.321) |
| F x SpouseFT x Child<4 | | | -0.152* | (0.025) |
| Constant | 11.218* | (0.022) | 3.192 | (0.454) |
| N | 1626 | | 1626 | |
| BIC | 1811.556 | | -24.439 | |
| F-test | | | 8.812*** | |
| Endogeneity(DWH) | 4.074* | | | |
| Hansen J stats p-value | 0.332 | | | |
| Adjusted $R^2$ | 0.515 | | 0.997 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001

## Table A2. First stage log hours regressions for models 2, 3 Table 2

| | Income models - 1st stage | | | |
|---|---|---|---|---|
| | Model 2 | | Model 3 | |
| | Full interactions | | Restricted interactions | |
| | Coef. | p-value | Coef. | p-value |
| Hours x Male | 0.000 | (1.000) | -0.025*** | (0.000) |
| Female | 9.308*** | (0.000) | 8.713*** | (0.000) |
| Service (10-20yrs) | 0.000 | (1.000) | -0.066*** | (0.001) |
| F x Service (10-20yrs) | -0.086* | (0.050) | 0.027 | (0.613) |
| Service (20-30yrs) | 0.000 | (1.000) | -0.071*** | (0.000) |
| F x Service (20-30yrs) | -0.075 | (0.094) | 0.054 | (0.339) |
| Service (>30yrs) | 0.000 | (1.000) | -0.075*** | (0.000) |
| F x Service (>30yrs) | -0.146** | (0.003) | -0.008 | (0.894) |
| NonWhite | -0.000 | (1.000) | 0.006 | (0.488) |
| F x NonWhite | -0.007 | (0.821) | -0.016 | (0.772) |
| Salaried | 0.000 | (1.000) | -0.301*** | (0.000) |
| F x Salaried | -0.428*** | (0.000) | | |
| ln List | -0.000 | (1.000) | -0.092** | (0.004) |
| F x ln List | -0.102 | (0.370) | -0.006 | (0.973) |
| List 2 | 0.000 | (1.000) | 0.017* | (0.023) |
| List 2 x Female | 0.036 | (0.195) | 0.019 | (0.632) |
| ln ListperGP | 0.000 | (1.000) | -0.122 | (0.740) |
| F x ln ListperGP | -0.430 | (0.536) | | |
| ListperGP 2 | -0.000 | (1.000) | 0.014 | (0.594) |
| F x ListperGP 2 | 0.042 | (0.377) | | |
| ln AgeSex | -0.000 | (1.000) | -0.053 | (0.128) |
| F x ln AgeSex | -0.099 | (0.442) | -0.005 | (0.976) |
| ln Needs | 0.000 | (1.000) | 0.016 | (0.723) |
| F x ln Needs | -0.156 | (0.397) | -0.327 | (0.176) |
| Dispensing | 0.000 | (1.000) | 0.009 | (0.533) |
| F x Dispensing | 0.045 | (0.107) | | |
| PMS | 0.000 | (1.000) | -0.002 | (0.870) |
| F x PMS | -0.011 | (0.593) | | |
| Optout | 0.000 | (1.000) | -0.010 | (0.591) |
| F x Optout | -0.027 | (0.359) | | |
| Amenities | 0.000 | (1.000) | -0.022 | (0.052) |
| F x Amenities | -0.034 | (0.109) | | |
| Rurality | 0.000 | (1.000) | -0.009 | (0.711) |
| F x Rurality | 0.040 | (0.607) | | |
| Traveltime | -0.000 | (1.000) | 0.000 | (0.870) |
| F x Traveltime | 0.000 | (0.878) | | |
| Femprop | 0.000 | (1.000) | 0.094** | (0.002) |
| F x Femprop | 0.148** | (0.008) | | |
| GOR6 | 0.000 | (1.000) | 0.005 | (0.388) |
| F x GOR6 | -0.084* | (0.018) | -0.097 | (0.053) |
| GOR7 | -0.000 | (1.000) | -0.002 | (0.813) |
| F x GOR7 | 0.064 | (0.086) | 0.034 | (0.515) |
| GOR8 | 0.000 | (1.000) | 0.003 | (0.589) |
| F x GOR8 | -0.047 | (0.131) | -0.071 | (0.130) |

| | | | | |
|---|---|---|---|---|
| GOR9 | 0.000 | (1.000) | 0.005 | (0.333) |
| F x GOR9 | -0.015 | (0.647) | -0.048 | (0.200) |
| F x Child<4 | -0.109** | (0.002) | -0.100* | (0.037) |
| F x Child4-18 | -0.215*** | (0.000) | -0.218*** | (0.000) |
| F x SpouseFT | -0.035* | (0.046) | -0.038 | (0.210) |
| F x SpouseFT x Child<4 | -0.162*** | (0.000) | -0.193** | (0.001) |
| Constant | -0.000 | (1.000) | 0.517 | (0.703) |
| N | 1902 | | 1902 | |
| BIC | -350.280 | | -291.046 | |
| F-test | 18.85*** | | 21.07*** | |
| Adjusted $R^2$ | 0.997 | | 0.997 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001

## Table A3. OLS income model (Table 3)

| | OLS, full set of female | |
|---|---|---|
| | Coef. | p-value |
| Hours x Female | 0.615*** | (0.000) |
| Hours x Male | 0.286*** | (0.000) |
| Female | -3.865 | (0.463) |
| Service (10-20yrs) | 0.0674 | (0.142) |
| F x Service (10-20yrs) | -0.193* | (0.023) |
| Service (20-30yrs) | 0.0724 | (0.097) |
| F x Service (20-30yrs) | -0.220** | (0.010) |
| Service (>30yrs) | 0.0630 | (0.156) |
| F x Service (>30yrs) | -0.109 | (0.218) |
| NonWhite | -0.0602 | (0.084) |
| F x NonWhite | 0.127 | (0.059) |
| Salaried | -0.426*** | (0.000) |
| F x Salaried | -0.229** | (0.009) |
| ln List | 0.105 | (0.347) |
| F x ln List | -0.482* | (0.047) |
| List 2 | -0.0134 | (0.620) |
| F x List 2 | 0.126* | (0.024) |
| ln ListperGP | -1.948*** | (0.000) |
| F x ln ListperGP | -0.195 | (0.885) |
| ListperGP 2 | 0.140*** | (0.000) |
| F x ListperGP 2 | 0.0140 | (0.880) |
| ln AgeSex | -0.00815 | (0.950) |
| F x ln AgeSex | -0.558* | (0.033) |
| ln Needs | -0.445* | (0.011) |
| F x ln Needs | 0.619 | (0.097) |
| Dispensing | 0.156*** | (0.000) |
| F x Dispensing | 0.0710 | (0.189) |
| PMS | 0.115*** | (0.000) |
| F x PMS | -0.0355 | (0.344) |
| Optout | 0.0912** | (0.002) |
| F x Optout | 0.0222 | (0.673) |
| Amenities | 0.0235 | (0.171) |
| F x Amenities | 0.0447 | (0.273) |
| Rurality | 0.116* | (0.031) |
| F x Rurality | -0.119 | (0.192) |
| Traveltime | -0.00143 | (0.072) |
| F x Traveltime | -0.000250 | (0.878) |
| Femprop | -0.116* | (0.021) |
| F x Femprop | 0.155 | (0.178) |
| GOR6 | -0.0307 | (0.237) |
| F x GOR6 | 0.0113 | (0.884) |
| GOR7 | -0.0425 | (0.437) |
| F x GOR7 | 0.199* | (0.012) |
| GOR8 | -0.0386 | (0.297) |

| | | |
|---|---|---|
| F x GOR8 | -0.00604 | (0.925) |
| GOR9 | -0.112*** | (0.000) |
| F x GOR9 | 0.0947 | (0.081) |
| Constant | 10.96*** | (0.000) |
| N | 1902 | |
| BIC | 1719.82 | |
| Adjusted $R^2$ | 0.521 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001

**Table A4.1. Second stage log income models for direct test with unisex variable**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | pooled | | model2 | | model3 | | nested | |
| Hours x Female | 0.954*** | (0.000) | 0.927*** | (0.000) | 0.954*** | (0.000) | 0.998*** | (0.000) |
| Female | -6.772 | (0.286) | -9.661 | (0.085) | -5.810** | (0.007) | -4.581*** | (0.000) |
| Hours x Male | 0.295** | (0.003) | 0.306*** | (0.000) | 0.304*** | (0.000) | 0.312*** | (0.000) |
| Unisex | -0.074 | (0.963) | 15.065** | (0.004) | 6.205** | (0.008) | 3.021 | (0.093) |
| Unisex x F | 4.739 | (0.057) | -6.691 | (0.524) | 0.074 | (0.993) | 4.926* | (0.014) |
| Hours x F x U | -0.611* | (0.020) | -0.612*** | (0.000) | -0.697*** | (0.000) | -0.688** | (0.002) |
| Hours x Male x U | 0.017 | (0.934) | 0.019 | (0.901) | 0.003 | (0.986) | -0.052 | (0.736) |
| List 3 | -0.008 | (0.801) | 0.047 | (0.744) | 0.088 | (0.576) | | |
| List 4 | 0.003 | (0.682) | -0.008 | (0.706) | -0.012 | (0.577) | | |
| ListperGP 3 | 0.006* | (0.017) | -0.016 | (0.496) | -0.015 | (0.550) | | |
| Service (10-20yrs) | 0.090 | (0.054) | 0.059 | (0.231) | 0.052 | (0.293) | | |
| Service (10-20yrs) x F | -0.141 | (0.132) | -0.159 | (0.064) | -0.147 | (0.072) | | |
| Service (10-20yrs) x U | | | 0.089 | (0.382) | 0.035 | (0.760) | | |
| Service (10-20yrs) x F x U | | | 0.129 | (0.435) | 0.112 | (0.481) | | |
| Service (20-30yrs) | 0.092* | (0.039) | 0.076 | (0.118) | 0.067 | (0.163) | 0.014 | (0.429) |
| Service (20-30yrs) x F | -0.177 | (0.067) | -0.215* | (0.014) | -0.198* | (0.015) | -0.077* | (0.037) |
| Service (20-30yrs) x U | | | -0.058 | (0.593) | -0.082 | (0.488) | | |
| Service (20-30yrs) x F x U | | | -0.059 | (0.638) | 0.138 | (0.358) | | |
| Service (>30yrs) | 0.076 | (0.113) | 0.053 | (0.289) | 0.042 | (0.386) | | |
| Service (>30yrs) x F | -0.055 | (0.581) | -0.116 | (0.204) | -0.092 | (0.296) | | |
| Service (>30yrs) x U | | | 0.068 | (0.477) | 0.063 | (0.525) | | |
| NonWhite | -0.057 | (0.148) | -0.017 | (0.642) | -0.018 | (0.609) | 0.019 | (0.589) |
| NonWhite x F | 0.158 | (0.056) | 0.097 | (0.196) | 0.092 | (0.215) | | |
| NonWhite x U | | | -0.195* | (0.045) | -0.169 | (0.074) | -0.159 | (0.050) |
| NonWhite x F x U | | | 0.032 | (0.847) | 0.042 | (0.796) | | |
| Salaried | -0.392*** | (0.000) | -0.445*** | (0.000) | -0.485*** | (0.000) | -0.469*** | (0.000) |
| Salaried x F | -0.103 | (0.447) | -0.068 | (0.524) | | | | |
| Salaried x U | | | 0.067 | (0.576) | | | | |
| Salaried x F x U | | | -0.706** | (0.001) | | | | |
| ln List | 0.064 | (0.650) | -0.110 | (0.746) | 0.032 | (0.941) | -0.314* | (0.025) |
| ln List x F | -0.007 | (0.928) | -0.375 | (0.382) | -0.209 | (0.527) | | |
| ln List x U | | | 0.443 | (0.087) | 0.493 | (0.069) | 0.564* | (0.018) |
| ln List x F x U | | | -0.476 | (0.551) | 0.176 | (0.823) | | |
| List 2 | | | -0.041 | (0.906) | -0.169 | (0.673) | 0.088** | (0.004) |
| List 2 x F | | | 0.091 | (0.330) | 0.056 | (0.446) | | |
| List 2 x F x U | | | -0.112 | (0.089) | -0.118 | (0.090) | -0.136* | (0.030) |
| List 2 x F x U | | | 0.203 | (0.440) | -0.104 | (0.714) | | |
| ln ListperGP | -0.921* | (0.029) | -3.848 | (0.262) | -3.683 | (0.324) | | |
| ln ListperGP x F | -0.024 | (0.839) | 0.837 | (0.528) | | | | |
| ln ListperGP x U | | | -2.390 | (0.086) | | | | |
| ln ListperGP x F x U | | | -0.297 | (0.923) | | | | |
| ListperGP 2 | | | 0.449 | (0.368) | 0.427 | (0.426) | | |
| ListperGP 2 x F | | | -0.054 | (0.556) | | | | |
| ListperGP 2 x F x U | | | 0.173 | (0.071) | | | | |
| ListperGP 2 x F x U | | | -0.044 | (0.835) | | | | |
| ln AgeSex | -0.182 | (0.327) | 0.138 | (0.308) | 0.115 | (0.393) | 0.152 | (0.218) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ln AgeSex x F | -0.505 | (0.198) | -0.705* | (0.011) | -0.628* | (0.013) | -0.740** | (0.001) |
| ln AgeSex x U | | | -0.594 | (0.074) | -0.731* | (0.036) | -0.712* | (0.045) |
| ln AgeSex x F x U | | | 0.094 | (0.918) | 1.187 | (0.211) | | |
| ln Needs | -0.307 | (0.286) | -0.279 | (0.112) | -0.221 | (0.207) | -0.041 | (0.786) |
| ln Needs x F | 0.758 | (0.203) | 0.586 | (0.137) | 0.360 | (0.329) | | |
| ln Needs x U | | | -1.486* | (0.015) | -1.266* | (0.026) | -0.518 | (0.144) |
| ln Needs x F x U | | | 3.647* | (0.041) | 0.888 | (0.622) | | |
| Dispensing | 0.156*** | (0.000) | 0.165*** | (0.000) | 0.176*** | (0.000) | 0.177*** | (0.000) |
| Dispensing x F | 0.063 | (0.335) | 0.066 | (0.234) | | | | |
| Dispensing x U | | | -0.150 | (0.083) | | | | |
| Dispensing x F x U | | | 0.443* | (0.047) | | | | |
| PMS | 0.125*** | (0.000) | 0.097*** | (0.000) | 0.101*** | (0.000) | 0.098*** | (0.000) |
| PMS x F | -0.001 | (0.987) | -0.012 | (0.768) | | | | |
| PMS x U | | | 0.054 | (0.273) | | | | |
| PMS x F x U | | | -0.116 | (0.356) | | | | |
| Optout | 0.106*** | (0.000) | 0.061 | (0.057) | 0.095*** | (0.000) | 0.099*** | (0.000) |
| Optout x F | -0.005 | (0.938) | 0.056 | (0.355) | | | | |
| Optout x U | | | 0.145* | (0.027) | | | | |
| Optout x F x U | | | -0.646** | (0.004) | | | | |
| Amenities | 0.012 | (0.561) | 0.019 | (0.293) | 0.041* | (0.023) | 0.048** | (0.008) |
| Amenities x F | 0.076 | (0.164) | 0.062 | (0.144) | | | | |
| Amenities x U | | | 0.002 | (0.966) | | | | |
| Amenities x F x U | | | 0.033 | (0.807) | | | | |
| Rurality | 0.134* | (0.023) | 0.061 | (0.340) | 0.028 | (0.579) | | |
| Rurality x F | -0.136 | (0.188) | -0.097 | (0.359) | | | | |
| Rurality x U | | | 0.114 | (0.379) | | | | |
| Traveltime | -0.001 | (0.089) | -0.001 | (0.131) | -0.002* | (0.022) | -0.002* | (0.025) |
| Traveltime x F | -0.001 | (0.536) | -0.001 | (0.549) | | | | |
| Traveltime x U | | | -0.003 | (0.185) | | | | |
| Traveltime x F x U | | | 0.005 | (0.334) | | | | |
| NonUK | -0.006 | (0.868) | | | | | | |
| NonUK x F | -0.102 | (0.212) | | | | | | |
| Senior | 0.001 | (0.962) | | | | | | |
| Senior x F | 0.018 | (0.777) | | | | | | |
| ln LISI | -0.018 | (0.552) | | | | | | |
| ln LISI x F | 0.004 | (0.952) | | | | | | |
| ln Noqual | -0.003 | (0.957) | | | | | | |
| ln Noqual x F | -0.033 | (0.762) | | | | | | |
| ln Nursing home | 0.010 | (0.416) | | | | | | |
| ln Nursing home x F | -0.039 | (0.215) | | | | | | |
| ln QOF points | 0.203 | (0.728) | | | | | | |
| ln QOF points x F | -0.222 | (0.783) | | | | | | |
| Extendedhours | 0.012 | (0.299) | | | | | | |
| Extendedhours x F | -0.001 | (0.983) | | | | | | |
| Housing | 0.007 | (0.746) | | | | | | |
| Housing x F | -0.014 | (0.762) | | | | | | |
| GOODschools | 0.019 | (0.427) | | | | | | |
| GOODschools x F | 0.007 | (0.889) | | | | | | |
| Competition | -0.002 | (0.759) | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Competition x F | 0.004 | (0.700) | | | | | | |
| ln MFFstaff | 0.280 | (0.394) | | | | | | |
| ln MFFstaff x F | -0.536 | (0.434) | | | | | | |
| ln MFFpremises | -0.048 | (0.277) | | | | | | |
| ln MFFpremises x F | -0.052 | (0.551) | | | | | | |
| GOR2 | 0.041 | (0.330) | | | | | | |
| GOR2 x F | 0.147 | (0.192) | | | | | | |
| GOR3 | 0.035 | (0.494) | | | | | | |
| GOR3 x F | 0.161 | (0.163) | | | | | | |
| GOR4 | 0.033 | (0.533) | | | | | | |
| GOR4 x F | 0.152 | (0.216) | | | | | | |
| GOR5 | 0.060 | (0.195) | | | | | | |
| GOR5 x F | 0.167 | (0.167) | | | | | | |
| GOR6 | 0.030 | (0.551) | -0.025 | (0.376) | -0.022 | (0.438) | | |
| GOR6 x F | 0.255 | (0.121) | 0.045 | (0.585) | 0.035 | (0.671) | | |
| GOR6 x U | | | 0.019 | (0.783) | -0.029 | (0.684) | | |
| GOR6 x F x U | | | 0.592* | (0.022) | 0.033 | (0.909) | | |
| GOR7 | 0.007 | (0.941) | 0.029 | (0.616) | 0.032 | (0.577) | 0.090* | (0.030) |
| GOR7 x F | 0.479* | (0.017) | 0.132 | (0.136) | 0.115 | (0.178) | | |
| GOR7 x U | | | -0.441*** | (0.001) | -0.415*** | (0.001) | -0.298*** | (0.001) |
| GOR7 x F x U | | | 0.752* | (0.014) | 0.256 | (0.374) | | |
| GOR8 | 0.012 | (0.845) | -0.012 | (0.729) | -0.009 | (0.800) | | |
| GOR8 x F | 0.226 | (0.163) | -0.003 | (0.971) | -0.015 | (0.831) | | |
| GOR8 x U | | | -0.234 | (0.060) | -0.226 | (0.066) | | |
| GOR8 x F x U | | | 0.416 | (0.090) | 0.039 | (0.864) | | |
| GOR9 | -0.052 | (0.276) | -0.088** | (0.001) | -0.088** | (0.001) | -0.067** | (0.004) |
| GOR9 x F | 0.260* | (0.029) | 0.111* | (0.043) | 0.096 | (0.074) | | |
| GOR9 x U | | | -0.099 | (0.355) | -0.116 | (0.326) | | |
| GOR9 x F x U | | | -0.061 | (0.843) | -0.320 | (0.266) | | |
| Constant | 6.479 | (0.153) | 13.917 | (0.079) | 13.169 | (0.129) | 2.372** | (0.008) |
| N | 1626 | | 1902 | | 1902 | | 1902 | |
| BIC | 1819.22 | | 2024.95 | | 1878.14 | | 1675.19 | |
| F-test | 7.57*** | | 16.73*** | | 18.94*** | | 22.82*** | |
| Endogeneity(DWH) | 3.464* | | 3.653* | | 4411.707*** | | 5.328** | |
| Hansen J stats p-value | 0.350 | | 0.000 | | 0.000 | | 0.118 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001

## Table A4.2. First stage log hours models for direct test with unisex variable

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | Pooled | | Model 2 | | Model 3 | | Nested | |
| Female | 7.150 | (0.027) | 16.410*** | (0.000) | 8.946*** | (0.000) | 7.295*** | (0.000) |
| Hours x Male | -0.000 | (1.000) | 0.000 | (0.894) | -0.023** | (0.007) | -0.027** | (0.005) |
| Unisex | -0.000 | (1.000) | 0.114 | (0.770) | 0.496 | (0.252) | -0.417 | (0.377) |
| Unisex x F | -6.085 | (0.000) | -16.274*** | (0.000) | -8.677*** | (0.001) | -6.693*** | (0.000) |
| Hours x F x U | 0.836 | (0.000) | 0.951*** | (0.000) | 0.920*** | (0.000) | 0.919*** | (0.000) |
| Hours x Male x U | 0.000 | (1.000) | -0.000 | (0.829) | -0.002 | (0.876) | 0.007 | (0.632) |
| List 3 | | | -0.031 | (0.599) | -0.057 | (0.433) | | |
| List 4 | | | 0.003 | (0.699) | 0.004 | (0.653) | | |
| ListperGP 3 | | | 0.004 | (0.798) | 0.013 | (0.479) | | |
| Service (10-20yrs) | -0.000 | (1.000) | -0.001 | (0.549) | -0.057** | (0.005) | | |
| Service (10-20yrs) x F | -0.080 | (0.085) | -0.074 | (0.161) | 0.026 | (0.634) | | |
| Service (10-20yrs) x U | | | 0.001 | (0.465) | -0.046 | (0.312) | | |
| Service (10-20yrs) x F x U | | | 0.049 | (0.588) | 0.001 | (0.992) | | |
| Service (20-30yrs) | -0.000 | (1.000) | -0.001 | (0.535) | -0.064** | (0.002) | -0.006 | (0.087) |
| Service (20-30yrs) x F | -0.096 | (0.048) | -0.061 | (0.299) | 0.055 | (0.340) | 0.033 | (0.223) |
| Service (20-30yrs) x U | | | 0.002 | (0.441) | -0.044 | (0.352) | | |
| Service (20-30yrs) x F x U | | | 0.032 | (0.493) | -0.032 | (0.553) | | |
| Service (>30yrs) | -0.000 | (1.000) | -0.001 | (0.567) | -0.062** | (0.002) | | |
| Service (>30yrs) x F | -0.143 | (0.010) | -0.123 | (0.054) | -0.008 | (0.898) | | |
| Service (>30yrs) x U | | | 0.002 | (0.420) | -0.048 | (0.289) | | |
| NonWhite | -0.000 | (1.000) | 0.001 | (0.576) | 0.008 | (0.388) | 0.003 | (0.885) |
| NonWhite x F | -0.047 | (0.208) | -0.007 | (0.906) | -0.016 | (0.778) | | |
| NonWhite x U | | | -0.001 | (0.578) | -0.006 | (0.705) | -0.005 | (0.833) |
| NonWhite x F x U | | | -0.108 | (0.153) | -0.066 | (0.433) | | |
| Salaried | -0.000 | (1.000) | -0.003 | (0.549) | -0.302*** | (0.000) | -0.287*** | (0.000) |
| Salaried x F | -0.434 | (0.000) | -0.430*** | (0.000) | | | | |
| Salaried x U | | | 0.003 | (0.573) | | | | |
| Salaried x F x U | | | 0.349*** | (0.000) | | | | |
| ln List | 0.000 | (1.000) | -0.138 | (0.413) | -0.454* | (0.028) | 0.028 | (0.763) |
| ln List x F | 0.130 | (0.000) | 0.334 | (0.234) | 0.351 | (0.106) | | |
| ln List x U | | | 0.022 | (0.719) | 0.168 | (0.062) | -0.019 | (0.850) |
| ln List x F x U | | | -0.331 | (0.401) | -0.350 | (0.365) | | |
| List 2 | | | 0.104 | (0.492) | 0.250 | (0.190) | -0.001 | (0.980) |
| List 2 x F | | | -0.059 | (0.343) | -0.058 | (0.222) | | |
| List 2 x U | | | -0.007 | (0.662) | -0.039 | (0.071) | -0.003 | (0.901) |
| List 2 x F x U | | | 0.013 | (0.914) | 0.057 | (0.662) | | |
| ln ListperGP | -0.000 | (1.000) | 0.525 | (0.793) | 1.385 | (0.582) | | |
| ln ListperGP x F | 0.127 | (0.011) | -2.175* | (0.041) | | | | |
| ln ListperGP x U | | | -0.044 | (0.680) | | | | |
| ln ListperGP x F x U | | | 4.170** | (0.004) | | | | |
| ListperGP 2 | | | -0.074 | (0.796) | -0.230 | (0.530) | | |
| ListperGP 2 x F | | | 0.157* | (0.034) | | | | |
| ListperGP 2 x U | | | 0.003 | (0.652) | | | | |
| ListperGP 2 x F x U | | | -0.287** | (0.004) | | | | |
| ln AgeSex | 0.000 | (1.000) | 0.007 | (0.479) | -0.031 | (0.348) | -0.073 | (0.160) |
| ln AgeSex x F | -0.033 | (0.864) | -0.133 | (0.423) | -0.038 | (0.815) | -0.033 | (0.833) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ln AgeSex x U | | | -0.004 | (0.627) | 0.047 | (0.365) | 0.072 | (0.280) |
| ln AgeSex x F x U | | | 0.726 | (0.085) | 0.146 | (0.671) | | |
| ln Needs | -0.000 | (1.000) | -0.007 | (0.358) | -0.008 | (0.859) | -0.075 | (0.389) |
| ln Needs x F | -0.591 | (0.070) | -0.349 | (0.163) | -0.452 | (0.060) | | |
| ln Needs x U | | | 0.007 | (0.411) | -0.141 | (0.092) | 0.077 | (0.436) |
| ln Needs x F x U | | | -1.344 | (0.077) | 0.495 | (0.407) | | |
| Dispensing | -0.000 | (1.000) | -0.000 | (0.613) | 0.003 | (0.850) | 0.003 | (0.804) |
| Dispensing x F | 0.032 | (0.299) | 0.029 | (0.433) | | | | |
| Dispensing x U | | | -0.001 | (0.586) | | | | |
| Dispensing x F x U | | | 0.175 | (0.077) | | | | |
| PMS | -0.000 | (1.000) | -0.000 | (0.885) | -0.002 | (0.844) | -0.007 | (0.435) |
| PMS x F | -0.011 | (0.627) | -0.006 | (0.807) | | | | |
| PMS x U | | | 0.001 | (0.433) | | | | |
| PMS x F x U | | | 0.089 | (0.095) | | | | |
| Optout | 0.000 | (1.000) | 0.000 | (0.700) | -0.007 | (0.702) | 0.002 | (0.909) |
| Optout x F | -0.006 | (0.842) | -0.018 | (0.669) | | | | |
| Optout x U | | | 0.000 | (0.905) | | | | |
| Optout x F x U | | | 0.117 | (0.066) | | | | |
| Amenities | -0.000 | (1.000) | -0.000 | (0.983) | -0.015 | (0.182) | -0.015 | (0.179) |
| Amenities x F | -0.033 | (0.201) | -0.027 | (0.375) | | | | |
| Amenities x U | | | 0.001 | (0.529) | | | | |
| Amenities x F x U | | | 0.044 | (0.477) | | | | |
| Rurality | 0.000 | (1.000) | 0.001 | (0.664) | 0.005 | (0.858) | | |
| Rurality x F | 0.092 | (0.258) | 0.080 | (0.252) | | | | |
| Rurality x U | | | -0.002 | (0.660) | | | | |
| Traveltime | 0.000 | (1.000) | 0.000 | (0.509) | 0.000 | (0.584) | 0.000 | (0.550) |
| Traveltime x F | 0.000 | (0.658) | 0.000 | (0.820) | | | | |
| Traveltime x U | | | -0.000 | (0.447) | | | | |
| Traveltime x F x U | | | 0.003 | (0.114) | | | | |
| GOR6 | -0.000 | (1.000) | -0.001 | (0.698) | 0.008 | (0.263) | | |
| GOR6 x F | -0.278 | (0.000) | -0.104* | (0.030) | -0.122* | (0.015) | | |
| GOR6 x U | | | 0.001 | (0.624) | -0.024 | (0.068) | | |
| GOR6 x F x U | | | -0.092 | (0.444) | 0.069 | (0.552) | | |
| GOR7 | -0.000 | (1.000) | -0.000 | (0.720) | 0.004 | (0.663) | 0.019 | (0.462) |
| GOR7 x F | -0.257 | (0.002) | 0.016 | (0.774) | -0.017 | (0.769) | | |
| GOR7 x U | | | -0.001 | (0.641) | -0.021 | (0.152) | -0.033 | (0.253) |
| GOR7 x F x U | | | -0.227* | (0.044) | 0.032 | (0.770) | | |
| GOR8 | -0.000 | (1.000) | -0.001 | (0.574) | 0.002 | (0.689) | | |
| GOR8 x F | -0.240 | (0.000) | -0.072 | (0.141) | -0.098* | (0.044) | | |
| GOR8 x U | | | -0.000 | (0.912) | -0.023 | (0.190) | | |
| GOR8 x F x U | | | -0.069 | (0.525) | 0.106 | (0.287) | | |
| GOR9 | -0.000 | (1.000) | -0.001 | (0.579) | 0.001 | (0.787) | -0.007 | (0.575) |
| GOR9 x F | -0.127 | (0.029) | -0.039 | (0.328) | -0.066 | (0.080) | | |
| GOR9 x U | | | 0.000 | (0.737) | 0.004 | (0.806) | | |
| GOR9 x F x U | | | -0.277 | (0.054) | 0.032 | (0.714) | | |
| Child<4 x F | -0.080 | (0.040) | -0.112* | (0.021) | -0.103* | (0.027) | -0.093* | (0.029) |
| Child4-18 x F | -0.159 | (0.000) | -0.211*** | (0.000) | -0.214*** | (0.000) | -0.204*** | (0.000) |
| SpouseFT x F | -0.027 | (0.164) | -0.024 | (0.403) | -0.024 | (0.408) | -0.032 | (0.259) |
| SpouseFT x Child<4 x F | -0.146 | (0.001) | -0.154* | (0.011) | -0.188** | (0.001) | -0.176** | (0.002) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NonUK | -0.000 | (1.000) | | | | | | |
| NonUK x F | 0.136 | (0.000) | | | | | | |
| Senior | 0.000 | (1.000) | | | | | | |
| Senior x F | 0.141 | (0.000) | | | | | | |
| ln LISI | 0.000 | (1.000) | | | | | | |
| ln LISI x F | -0.019 | (0.542) | | | | | | |
| ln Noqual | 0.000 | (1.000) | | | | | | |
| ln Noqual x F | 0.111 | (0.069) | | | | | | |
| ln Nursing home | 0.000 | (1.000) | | | | | | |
| ln Nursing home x F | 0.027 | (0.031) | | | | | | |
| ln QOF points | -0.000 | (1.000) | | | | | | |
| ln QOF points x F | 0.344 | (0.388) | | | | | | |
| Extendedhours | -0.000 | (1.000) | | | | | | |
| Extendedhours x F | 0.006 | (0.645) | | | | | | |
| Housing | 0.000 | (1.000) | | | | | | |
| Housing x F | 0.020 | (0.421) | | | | | | |
| GOODschools | -0.000 | (1.000) | | | | | | |
| GOODschools x F | -0.026 | (0.337) | | | | | | |
| Competition | 0.000 | (1.000) | | | | | | |
| Competition x F | 0.009 | (0.112) | | | | | | |
| ln MFFstaff | -0.000 | (1.000) | | | | | | |
| ln MFFstaff x F | 0.695 | (0.044) | | | | | | |
| ln MFFpremises | 0.000 | (1.000) | | | | | | |
| ln MFFpremises x F | 0.026 | (0.550) | | | | | | |
| GOR2 | -0.000 | (1.000) | | | | | | |
| GOR2x F | -0.121 | (0.030) | | | | | | |
| GOR3 | -0.000 | (1.000) | | | | | | |
| GOR3 x F | -0.086 | (0.139) | | | | | | |
| GOR4 | -0.000 | (1.000) | | | | | | |
| GOR4 x F | -0.088 | (0.159) | | | | | | |
| GOR5 | -0.000 | (1.000) | | | | | | |
| GOR5 x F | -0.096 | (0.123) | | | | | | |
| Constant | 0.000 | (1.000) | -1.146 | (0.801) | -2.122 | (0.711) | 0.574 | (0.146) |
| N | 1626 | | 1902 | | 1902 | | 1902 | |
| BIC | -708.334 | | -165.381 | | -226.222 | | -427.024 | |
| F-test | 7.573*** | | 16.729*** | | 18.941*** | | 22.820*** | |
| Adjusted $R^2$ | 0.997 | | 0.997 | | 0.997 | | 0.997 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001.

**Table A5.1. Second stage log income models for direct tests with female senior GP, proportion of female GPs**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hours x Female | 0.918*** | (0.000) | 0.911*** | (0.000) | 0.936*** | (0.000) | 0.933*** | (0.000) | 0.927*** | (0.000) |
| Hours x Male | 0.282*** | (0.000) | 0.278*** | (0.000) | 0.283*** | (0.000) | 0.284*** | (0.000) | 0.281*** | (0.000) |
| Female (F) | -6.292** | (0.002) | -6.443** | (0.002) | -6.755** | (0.001) | -6.911** | (0.001) | -7.107*** | (0.001) |
| Femprop | -0.117* | (0.023) | | | | | -0.138* | (0.010) | | |
| F x Femprop | 0.137 | (0.259) | | | | | 0.136 | (0.283) | | |
| Service (10-20yrs) | 0.051 | (0.270) | 0.045 | (0.339) | 0.055 | (0.241) | 0.050 | (0.281) | 0.044 | (0.342) |
| F x Service (10-20yrs) | -0.145 | (0.072) | -0.141 | (0.073) | -0.145 | (0.072) | -0.140 | (0.084) | -0.139 | (0.078) |
| Service (20-30yrs) | 0.057 | (0.193) | 0.049 | (0.259) | 0.054 | (0.209) | 0.050 | (0.250) | 0.043 | (0.319) |
| F x Service (20-30yrs) | -0.180* | (0.022) | -0.168* | (0.027) | -0.183* | (0.021) | -0.177* | (0.025) | -0.168* | (0.028) |
| Service (>30yrs) | 0.046 | (0.298) | 0.038 | (0.387) | 0.027 | (0.547) | 0.026 | (0.561) | 0.020 | (0.662) |
| F x Service (>30yrs) | -0.075 | (0.371) | -0.066 | (0.423) | -0.066 | (0.429) | -0.064 | (0.447) | -0.059 | (0.477) |
| NonWhite | -0.059 | (0.088) | -0.062 | (0.075) | -0.047 | (0.186) | -0.055 | (0.114) | -0.057 | (0.099) |
| F x NonWhite | 0.121 | (0.086) | 0.126 | (0.074) | 0.113 | (0.116) | 0.122 | (0.090) | 0.126 | (0.079) |
| Salaried | -0.482*** | (0.000) | -0.491*** | (0.000) | -0.474*** | (0.000) | -0.472*** | (0.000) | -0.479*** | (0.000) |
| ln List | 0.344 | (0.580) | 0.368 | (0.563) | 0.445 | (0.506) | 0.376 | (0.582) | 0.406 | (0.558) |
| F x ln List | -0.414 | (0.075) | -0.467 | (0.058) | -0.304 | (0.151) | -0.369 | (0.096) | -0.430 | (0.069) |
| List 2 | -0.170 | (0.746) | -0.196 | (0.714) | -0.258 | (0.647) | -0.145 | (0.800) | -0.183 | (0.754) |
| F x List 2 | 0.105 | (0.051) | 0.116* | (0.040) | 0.080 | (0.105) | 0.094 | (0.065) | 0.107* | (0.048) |
| ListperGP | -4.822 | (0.259) | -4.871 | (0.238) | -5.191 | (0.169) | -5.026 | (0.164) | -4.986 | (0.151) |
| ListperGP sq | 0.545 | (0.368) | 0.558 | (0.340) | 0.619 | (0.255) | 0.594 | (0.256) | 0.595 | (0.236) |
| ln AgeSex | -0.031 | (0.811) | -0.039 | (0.765) | -0.034 | (0.802) | -0.030 | (0.822) | -0.039 | (0.768) |
| F x ln AgeSex | -0.452 | (0.082) | -0.419 | (0.100) | -0.495 | (0.051) | -0.496 | (0.057) | -0.465 | (0.068) |
| ln Needs | -0.384* | (0.026) | -0.416* | (0.020) | -0.371* | (0.035) | -0.426* | (0.013) | -0.452* | (0.011) |
| F x ln Needs | 0.478 | (0.182) | 0.534 | (0.142) | 0.547 | (0.133) | 0.594 | (0.101) | 0.659 | (0.073) |
| Dispensing | 0.177*** | (0.000) | 0.178*** | (0.000) | 0.182*** | (0.000) | 0.181*** | (0.000) | 0.183*** | (0.000) |
| PMS | 0.104*** | (0.000) | 0.104*** | (0.000) | 0.110*** | (0.000) | 0.107*** | (0.000) | 0.106*** | (0.000) |
| Optout | 0.100*** | (0.000) | 0.099*** | (0.000) | 0.098*** | (0.000) | 0.097*** | (0.000) | 0.096*** | (0.000) |
| Amenities | 0.044* | (0.013) | 0.042* | (0.016) | 0.043* | (0.020) | 0.043* | (0.017) | 0.041* | (0.024) |
| Rurality | 0.051 | (0.287) | 0.045 | (0.367) | 0.060 | (0.197) | 0.059 | (0.214) | 0.051 | (0.298) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Traveltime | -0.001 | (0.064) | -0.001 | (0.061) | -0.001 | (0.063) | -0.001 | (0.077) | -0.001 | (0.070) |
| GOR6 | -0.027 | (0.291) | -0.032 | (0.217) | -0.028 | (0.283) | -0.033 | (0.199) | -0.038 | (0.146) |
| F x GOR6 | 0.026 | (0.745) | 0.027 | (0.732) | 0.034 | (0.663) | 0.039 | (0.619) | 0.040 | (0.603) |
| GOR7 | -0.035 | (0.511) | -0.035 | (0.505) | -0.046 | (0.380) | -0.042 | (0.429) | -0.042 | (0.420) |
| F x GOR7 | 0.163* | (0.039) | 0.179* | (0.026) | 0.179* | (0.022) | 0.174* | (0.027) | 0.193* | (0.016) |
| GOR8 | -0.036 | (0.328) | -0.040 | (0.294) | -0.037 | (0.320) | -0.042 | (0.246) | -0.045 | (0.229) |
| F x GOR8 | -0.001 | (0.984) | 0.004 | (0.949) | 0.011 | (0.876) | 0.015 | (0.826) | 0.024 | (0.737) |
| GOR9 | -0.109*** | (0.000) | -0.110*** | (0.000) | -0.118*** | (0.000) | -0.119*** | (0.000) | -0.119*** | (0.000) |
| F x GOR9 | 0.093 | (0.079) | 0.093 | (0.077) | 0.113* | (0.030) | 0.112* | (0.032) | 0.114* | (0.027) |
| List 3 | 0.036 | (0.845) | 0.046 | (0.807) | 0.065 | (0.745) | 0.017 | (0.931) | 0.032 | (0.874) |
| List 4 | -0.002 | (0.930) | -0.003 | (0.893) | -0.005 | (0.840) | 0.001 | (0.963) | -0.001 | (0.979) |
| ListperGP 3 | -0.019 | (0.504) | -0.020 | (0.468) | -0.023 | (0.370) | -0.022 | (0.377) | -0.023 | (0.348) |
| PropFem1/4to1/2 | | | -0.062** | (0.001) | | | | | -0.065*** | (0.001) |
| PropFem1/2to3/4 | | | -0.073 | (0.057) | | | | | -0.089* | (0.028) |
| PropFem>3/4 | | | -0.144 | (0.051) | | | | | -0.154* | (0.039) |
| F x PropFem1/4to1/2 | | | -0.019 | (0.774) | | | | | -0.013 | (0.847) |
| F x PropFem1/2to3/4 | | | 0.071 | (0.360) | | | | | 0.094 | (0.247) |
| F x PropFem>3/4 | | | 0.052 | (0.678) | | | | | 0.035 | (0.779) |
| NonSenior x FemSenior | | | | | 0.002 | (0.926) | 0.024 | (0.380) | 0.027 | (0.328) |
| F x NonSenior x FemSenior | | | | | -0.010 | (0.842) | -0.032 | (0.556) | -0.038 | (0.472) |
| Senior | | | | | 0.037 | (0.155) | 0.040 | (0.124) | 0.040 | (0.122) |
| F x Senior | | | | | -0.004 | (0.936) | -0.006 | (0.905) | -0.001 | (0.982) |
| Constant | 17.417 | (0.082) | 17.628 | (0.068) | 17.668* | (0.041) | 17.624* | (0.033) | 17.580* | (0.027) |
| N | 1902 | | 1902 | | 1893 | | 1893 | | 1893 | |
| BIC | 1743.918 | | 1758.657 | | 1744.144 | | 1752.874 | | 1766.533 | |
| F-test | 20.98*** | | 21.12*** | | 20.19*** | | 20.07*** | | 20.19*** | |
| Endogeneity(DWH) | 4.686** | | 4.600** | | 4.529** | | 4.487** | | 4.388** | |
| Hansen J stats p-value | 0.164 | | 0.136 | | 0.205 | | 0.198 | | 0.182 | |

p-values in parentheses, * $p<0.05$ ** $p<0.001$ *** $p<0.001$

**Table A5.2. First stage log hours models for direct tests with female senior GP, proportion of female GPs**

| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hours x Female | -0.025*** | (0.000) | -0.025*** | (0.001) | -0.022** | (0.001) | -0.022** | (0.001) | -0.021** | (0.002) |
| Hours x Male | 8.514*** | (0.000) | 8.556*** | (0.000) | 8.962*** | (0.000) | 8.861*** | (0.000) | 8.906*** | (0.000) |
| Female | 0.040* | (0.010) | | | | | 0.033* | (0.026) | | |
| Femprop | 0.143 | (0.077) | | | | | 0.090 | (0.276) | | |
| F x Femprop | -0.066*** | (0.001) | -0.069*** | (0.000) | -0.068*** | (0.000) | -0.064*** | (0.001) | -0.066*** | (0.001) |
| Service (10-20yrs) | 0.024 | (0.647) | 0.033 | (0.531) | 0.027 | (0.610) | 0.019 | (0.725) | 0.026 | (0.625) |
| F x Service (10-20yrs) | -0.072*** | (0.000) | -0.074*** | (0.000) | -0.072*** | (0.000) | -0.068*** | (0.000) | -0.070*** | (0.000) |
| Service (20-30yrs) | 0.053 | (0.349) | 0.056 | (0.314) | 0.035 | (0.542) | 0.028 | (0.628) | 0.030 | (0.591) |
| F x Service (20-30yrs) | -0.074*** | (0.000) | -0.076*** | (0.000) | -0.067*** | (0.001) | -0.064*** | (0.001) | -0.066*** | (0.001) |
| Service (>30yrs) | -0.017 | (0.784) | -0.018 | (0.779) | -0.066 | (0.299) | -0.078 | (0.220) | -0.079 | (0.215) |
| F x Service (>30yrs) | 0.003 | (0.688) | 0.003 | (0.669) | 0.001 | (0.928) | 0.003 | (0.673) | 0.003 | (0.634) |
| NonWhite | -0.012 | (0.835) | -0.011 | (0.841) | -0.007 | (0.898) | -0.008 | (0.881) | -0.008 | (0.882) |
| F x NonWhite | -0.295*** | (0.000) | -0.302*** | (0.000) | -0.300*** | (0.000) | -0.291*** | (0.000) | -0.297*** | (0.000) |
| Salaried | -0.397* | (0.031) | -0.425* | (0.020) | -0.371* | (0.035) | -0.348* | (0.046) | -0.373* | (0.032) |
| ln List | -0.015 | (0.928) | 0.030 | (0.858) | -0.109 | (0.471) | -0.051 | (0.736) | -0.006 | (0.970) |
| F x ln List | 0.244 | (0.128) | 0.283 | (0.076) | 0.244 | (0.116) | 0.217 | (0.161) | 0.252 | (0.101) |
| List 2 | 0.022 | (0.564) | 0.011 | (0.776) | 0.048 | (0.169) | 0.037 | (0.293) | 0.026 | (0.465) |
| F x List 2 | 1.109 | (0.657) | 0.792 | (0.745) | 1.380 | (0.565) | 1.243 | (0.621) | 0.928 | (0.711) |
| ListperGP | -0.165 | (0.654) | -0.123 | (0.731) | -0.219 | (0.537) | -0.204 | (0.585) | -0.163 | (0.662) |
| ListperGP sq | -0.043 | (0.178) | -0.031 | (0.331) | -0.034 | (0.278) | -0.035 | (0.247) | -0.024 | (0.423) |
| ln AgeSex | 0.006 | (0.973) | 0.000 | (0.999) | 0.071 | (0.673) | 0.100 | (0.568) | 0.097 | (0.578) |
| F x ln AgeSex | -0.014 | (0.754) | -0.031 | (0.484) | -0.040 | (0.363) | -0.021 | (0.622) | -0.036 | (0.393) |
| ln Needs | -0.298 | (0.213) | -0.306 | (0.210) | -0.379 | (0.106) | -0.385 | (0.101) | -0.397 | (0.097) |
| F x ln Needs | 0.010 | (0.495) | 0.008 | (0.571) | 0.003 | (0.837) | 0.005 | (0.704) | 0.004 | (0.770) |
| Dispensing | -0.002 | (0.836) | -0.002 | (0.818) | -0.006 | (0.543) | -0.005 | (0.623) | -0.005 | (0.605) |
| PMS | -0.011 | (0.546) | -0.010 | (0.557) | 0.000 | (0.991) | -0.002 | (0.914) | -0.002 | (0.925) |
| Optout | -0.020 | (0.071) | -0.017 | (0.133) | -0.023* | (0.041) | -0.022* | (0.045) | -0.019 | (0.085) |
| Amenities | -0.003 | (0.904) | 0.001 | (0.967) | -0.002 | (0.937) | 0.002 | (0.932) | 0.007 | (0.797) |
| Rurality | 0.000 | (0.815) | 0.000 | (0.817) | 0.000 | (0.907) | 0.000 | (0.995) | -0.000 | (0.985) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Traveltime | 0.001 | (0.893) | -0.001 | (0.894) | 0.001 | (0.890) | 0.002 | (0.747) | 0.000 | (0.935) |
| GOR6 | -0.097 | (0.053) | -0.094 | (0.059) | -0.093 | (0.059) | -0.099* | (0.043) | -0.098* | (0.045) |
| F x GOR6 | -0.002 | (0.842) | -0.004 | (0.641) | -0.000 | (0.960) | -0.000 | (0.966) | -0.003 | (0.752) |
| GOR7 | 0.027 | (0.612) | 0.024 | (0.658) | 0.019 | (0.712) | 0.013 | (0.804) | 0.008 | (0.876) |
| F x GOR7 | -0.001 | (0.822) | -0.005 | (0.397) | -0.003 | (0.619) | -0.001 | (0.840) | -0.004 | (0.426) |
| GOR8 | -0.070 | (0.136) | -0.074 | (0.124) | -0.074 | (0.106) | -0.080 | (0.082) | -0.084 | (0.072) |
| F x GOR8 | 0.004 | (0.485) | 0.003 | (0.598) | 0.002 | (0.698) | 0.002 | (0.633) | 0.001 | (0.764) |
| GOR9 | -0.045 | (0.228) | -0.048 | (0.203) | -0.041 | (0.267) | -0.041 | (0.268) | -0.044 | (0.232) |
| F x GOR9 | -0.062 | (0.294) | -0.078 | (0.184) | -0.065 | (0.257) | -0.054 | (0.341) | -0.070 | (0.221) |
| List 3 | 0.005 | (0.511) | 0.007 | (0.344) | 0.006 | (0.457) | 0.004 | (0.569) | 0.006 | (0.393) |
| List 4 | 0.009 | (0.634) | 0.007 | (0.702) | 0.012 | (0.502) | 0.011 | (0.543) | 0.009 | (0.607) |
| ListperGP 3 | -0.101* | (0.035) | -0.101* | (0.031) | -0.095* | (0.049) | -0.098* | (0.044) | -0.098* | (0.040) |
| PropFem1/4to1/2 | -0.221*** | (0.000) | -0.219*** | (0.000) | -0.208*** | (0.000) | -0.212*** | (0.000) | -0.210*** | (0.000) |
| PropFem1/2to3/4 | -0.036 | (0.233) | -0.039 | (0.184) | -0.031 | (0.296) | -0.030 | (0.321) | -0.033 | (0.264) |
| PropFem>3/4 | -0.195** | (0.001) | -0.195** | (0.001) | -0.195** | (0.001) | -0.196** | (0.001) | -0.196** | (0.001) |
| F x PropFem1/4to1/2 | | | 0.002 | (0.539) | | | | | 0.001 | (0.738) |
| F x PropFem1/2to3/4 | | | 0.016* | (0.038) | | | | | 0.012 | (0.097) |
| F x PropFem>3/4 | | | 0.117* | (0.015) | | | | | 0.111* | (0.018) |
| NonSenior x FemSenior | | | 0.030 | (0.561) | | | | | 0.019 | (0.708) |
| F x NonSenior x FemSenior | | | 0.028 | (0.609) | | | | | 0.009 | (0.877) |
| Senior | | | 0.063 | (0.476) | | | | | 0.031 | (0.720) |
| F x Senior | | | | | 0.003 | (0.586) | -0.002 | (0.748) | -0.002 | (0.681) |
| Constant | | | | | 0.052 | (0.120) | 0.040 | (0.264) | 0.040 | (0.259) |
| N | 1902 | | 1902 | | 1893 | | 1893 | | 1893 | |
| BIC | -270.888 | | -247.726 | | -312.758 | | -305.315 | | -283.888 | |
| F-test | 20.98*** | | 21.12*** | | 20.19*** | | 20.07*** | | 20.19*** | |
| Adjusted $R^2$ | 0.997 | | 0.997 | | 0.997 | | 0.997 | | 0.997 | |

p-values in parentheses, * p<0.05 ** p<0.001 *** p<0.001