THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2002/07

Estimating and Explaining Differences in Income Related Inequality in Health across General Practices

by

Mark Dusheiko, Hugh Gravelle and Stephen Campbell

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

Estimating and explaining differences in income related inequality in health across general practices

Mark Dusheiko[a] **Hugh Gravelle**[a,*] **Stephen Campbell**[b]

[a] *National Primary Care Research and Development Centre, Centre for Health Economics, University of York, UK*
[b] *National Primary Care Research and Development Centre, University of Manchester, UK*

Abstract. We use data on individual patients in general practices to examine whether income related inequality in self reported health differs across general practices and whether such differences are explained by characteristics of the practices. We allow for the simultaneous determination of health and income by instrumenting income. We also allow for item non response for the income question by a two stage selection model. We find that item non response has little effect on the estimated relationship between income and health but that allowing for simultaneity doubles the estimated effect of income on health. We show that there are significant differences in the effect of income on health across practices and that these differences are related to the number of patients per GP, a measure of practice prescribing quality, and the provision of out of hours services.

Keywords: Health. Income. Inequality. Primary care.

JEL Nos: I12, H40

[*] Correspondence to Hugh Gravelle, National Primary Care Research and Development Centre, Centre for Health Economics, Alcuin Block D, University of York, York YO10 5DD.

1. Introduction

One of the aims of government policy in the British National Health Service is to reduce differences between the health of rich and poor. The NHS is a publicly provided system funded almost entirely from taxation so that, to the extent that the NHS can affect health, reductions in inequality will require changes in the organisation and delivery of services, rather than changes in methods of financing to alter patient use.

Ninety per cent of patient contacts with the NHS are made via primary care. Patients are registered with a general practice and their general practitioners (GPs) act as gatekeepers, controlling non emergency access to the rest of the NHS. Most GPs are independent contractors, rather than employees. Even with recent attempts to introduce greater regulation, GPs have considerable freedom in the services they choose to provide to their patients and in the way they organise their practices to do so. Hence it is of interest to examine whether practice policies and organisation have any effect on inequalities in health between rich and poor.

The extent of income related inequality in health depends both on the effect of income on health and on the distribution of income [1,2]. In this paper we concentrate on the effect of income on health and consider whether it varies across practices and if so whether it is related to characteristics of practices, especially those which may be amenable to policy. To our knowledge this is the first attempt to test whether general practices have any effect on the relationship between income and health.

The basic approach is to estimate a regression of health on income, personal characteristics such as age, gender, and ethnicity and on the characteristics of the general practice to which the patient belongs. A positive coefficient on income indicates that there is income related inequality in health. To investigate if this differs across practices we interact income with practice dummies and test whether constraining all income coefficients to be equal leads to a significant reduction in the performance of the regression. We then investigate whether differences in the effect of income across practices are explained by characteristics of practices, such as the number of GPs or various indicators of practice quality. We do so in two ways. In the first method we estimate an individual patient level regression of health which includes practice dummy variables interacted with patient income, to obtain estimated income-health slope coefficients for each practice. Then we regress the estimated income slope coefficients for practices on practice characteristics. The second method is more direct: we regress individual health on individual characteristics, income, and the interaction of income with practice characteristics.

We also address two issues which do not appear to have been previously considered in the literature on income related inequality in health [3]. The first is selective non response which has two potential damaging consequences. First, attempts to increase the effective sample size by estimating income for non responders will yield biased income estimates if non response is related to income. Second, even if analysis is restricted to those who report income, the estimated effect of income on health will be biased if non-response is related to health because the same unobserved factors influence both the response to the income question and health.

The second issue is simultaneity: income affects health and health affects income. The standard procedure in analysis of income related inequality is to regress health on income but this will yield biased estimates of the effect of income on health. The estimate may be useful if one is only interested in measuring the overall correlation of income and health. For policy it is useful to know how much of the correlation is due to the effect of income on health and how much to the effect of health on income since different types of policy are required to change the two relationships.

2. Data

2.1 Patient and practice characteristics

The General Practice Assessment Survey (GPAS) (*www.gpas.co.uk*) asks patients about their use of general practice, their views on its accessibility and quality of care. We used an augmented version of GPAS, with additional questions on income, employment status and various aspects of health. The sample was selected by multistage stratification [4]. At the final stage approximately 200 adult patients were randomly selected from the lists of 60 practices. The sample was not self-weighting because the probability of an individual being selected depended on the size of the practice. We have not used sampling weights which are less efficient and no more unbiased in a regression model intending to determine the causal relationship between income and health [5]. The practices are in six Health Authorities and we include Health Authority dummy variables in the regressions as fixed effects to capture, inter alia, any survey design effects. We also allow for the clustering of errors within practices by using robust standard errors [6].

There were 4462 completed questionnaires, giving an overall response rate of 37%. There was a slight overrepresentation of females (59%) compared with practice populations and those over the age of 65 (27% actual against an expected 19%). Since the regression analysis conditions on observable characteristics of the sample respondents, the representativeness of the sample with respect to observable characteristics is not an important issue.

Some 3477 respondents completed all items on the questionnaire and a further 2539 completed all items except the income question (an item non response rate of 31%). After estimating income we therefore had a sample for analysis of the effects of practice characteristics of 3477. The variables are summarised in Table 1 for the full estimation sample.

Data on the characteristics of practices were obtained from the QUASAR study of practice quality [4] and from the Department of Health's General Medical Statistics database.

2.2 Health

The health measure used in the analysis is based on the SF-6D questionnaire included in GPAS. It covers six dimensions of health: physical functioning, role limitation, social functioning, pain, mental health, and vitality. Each dimension has between two and six levels. Weights were applied to responses to construct a single health measure [7] with 1 corresponding to the best possible health state and 0 to the worst.

3. Selection and simultaneity

3.1 Methods

We need to deal with two potential sources of bias in estimating the effect of income on health: selection bias arising from non response to the income question and simultaneous equation bias from the joint determination of health and income. We use a procedure suggested by Angrist [8] which is a combination of the Heckman two step selection correction for selection bias in the equation used to estimate income and two stage least squares estimation of the structural health equation to remove simultaneity bias and provide consistent estimates of the marginal effect of income on health.

We first estimate a selection equation for income non response using a probit regression using the sample of 3477 to obtain the inverse Mills ratio. We then use the sample of 2539 patients who completed all items to estimate income. Finally we estimate the health equation for the sample of 3477 individuals using estimated income and including the Mills ratio. To ensure the identification of the health equation we exclude from it two variables correlated with income but not directly correlated with health: the number of cars owned and the type of accommodation. To ensure that the Mills ratio from the selection equation is identified we exclude from the income and health equations some of the variables in the selection equation. The excluded variables were the length of time the patient had been with the practice and how convenient they found the location of the practice.

From Angrist [9] we know that when the instruments used in the income equation to identify the health equation are also in the sample selection (response) rule, the instruments may not be valid. The instruments will be independent of the error term in the health equation for the selected sample of responders to the income question if and only if selection status (i.e. report income or not) is independent of the errors in the health regression. To overcome this problem, we follow Angrist and include the inverse Mills ratio in the health equation. Since the Mills ratio reflects the propensity to report income its inclusion in the health equation will allow for any correlation of the error term and non-response.

We perform a variety of tests to check that we have good instruments, correlated with income but not affecting health directly, to identify the health equation. The instruments should not be correlated with the error term in the second stage health regression [10].

3.2 Results

3.2.1 Implications of selection and simultaneity

To examine the implications of selection and simultaneity we estimate four health equations (see Table 2): not allowing for selection or simultaneity, allowing for selection, allowing for simultaneity, and allowing for both selection and simultaneity. To separate out the effects of simultaneity and selection from those of sample size we estimate the health equation in all four regressions using the sample of individuals who responded to all questions. In the next section where we seek to examine the variation in the effect of income across practices we allow for both simultaneity and selection and so can increase the sample size by using predicted income thereby including patients who did not respond to the income question.

The explanatory variables include both individual characteristics (gender, age, ethnicity, marital status, number of children, smoking behaviour, household income) and characteristics of the practice to which the patient belongs, such as training status, number of GPs, patients per GP. Since there was a reported practice disability score for only 55 of the 60 practices the sample is the 2340 individuals who reported all items and belonged to these practices.

In all models the log of household income has a highly significant positive effect on self-reported health status. Comparing model 2 with model 1 we see that allowing for selective response to the income question has only a trivial effect on the coefficients and the inverse Mills ratio is not significant. Estimates using the propensity score (predicted values of the latent propensity to respond) and powers of the propensity produce virtually identical results.

Model 3 allows for potential endogeneity in the relationship between health and income by estimating the health equation using 2SLS, instrumenting income with the variables for the number of cars available to the responder and the type of accommodation, in terms of ownership, which they occupy. The augmented Hausman test shows that allowing for endogeneity leads to a significant increase in the magnitude of the effect of income on health. The estimated effect of log income doubles.

The final model allows for both endogeneity and selection bias in the health equation and the effect of income on health is the largest of the four models. The elasticity of health with respect to income is 0.0889. The increase in the income coefficient when selection is allowed for in addition to simultaneity (compare models 4 and 3) is much larger than the increase when only selection is allowed for (compare models 2 and 1), suggesting that the effects of simultaneity and selection are multiplicative. Moreover the inverse Mills ratio is now significant at the 5% level suggesting that allowing only for simultaneity will lead to biased estimates of the effect of income in the presence of income non response.

The estimated coefficients on the other variables appear plausible and reasonably stable across the different model specifications, though the magnitude of the effects of gender and smoking behaviour are altered by allowing for simultaneity and selection. The effect of gender is negative and significant, with a large effect in the final model. Women report 4% worse health than men. There is a nonlinear relationship between self-assessed health and age. Health deteriorates with age, but at a decreasing rate up until age 65, when the rate of decline increases. No significant effect of race was found.

The coefficients on marital status and the number of children capture not only any direct effect of family composition on health but also the effect via the implied change in equivalised income. Thus the fact that compared to the reference married state being single, divorced or widowed appear to increase health may be due either to their direct effect on health or to the fact that they may yield higher equivalised income. The positive coefficients on children seems to suggest that having children has a direct beneficial effect on health, since having more children reduces equivalised income. However, the family composition effects are not our main interest and they are generally not significant.

Not having smoked for more than a year is associated with better self-assessed health. The magnitude of the coefficient declines as we allow for endogeneity and selection bias. In the final model it is significant only at the at the 10% level.

The coefficients on practice characteristics suggest some association between patient health and practice organisation. Practices with better disability access scores and the proportion of GPs in the practice who have accredited training status were both found to be positively correlated with health. There was a non-linear relationship between the number of WTE GPs in the practice and health which is significant at the 10% level.

3.2.2 Validity of the instruments

We performed a number of tests to check that the variables (cars in the household, accommodation types) used to identify the effect of income in the health regression were reliable instruments in that they were correlated with income and uncorrelated with health. Both were statistically significant at the 1% level in a regression of income on all exogenous variables and the instruments. We tested whether the instruments had a direct effect on health by including one instrument at a time as an explanatory variable in the health regression while identifying income with the remaining instrumental variable. The results show that neither the number of cars ($F(2, 54) = 0.15$, $Prob > F = 0.863$) nor accommodation type ($F(3, 54) = 1.17$, $Prob > F = 0.3312$) had a significant direct effect on health status when instrumented income is included. The estimated coefficients on income was 0.081 when the number of cars was the instrument and 0.068 when accommodation type was the instrument).

Finally, in order to obtain consistent IV estimates of the income effect, we also require the instruments to be uncorrelated with the error term in the health equation. To test this assumption we followed Blundell and Smith [11] and regressed the residuals from the augmented Hausman regression on all the exogenous and instrumental variables in the model. A Chi squared statistic was calculated as the product of the R-squared from the regression and the number of observations. The statistic was compared with the critical value from the chi-squared distribution with degrees of freedom equal to the number of instruments. We were unable to reject the null hypothesis of no significant correlation between the instruments and the residual from the augmented Hausmann regression at the 5% significance level (n*R-squared = 4.68 < Chi2(5, 0.95) = 11.07). This conclusion is supported by a failure to reject the null hypothesis that the coefficients on the outside instruments are jointly equal to zero (F(5, 54) = 0.71, P-value = 0.6221).

4 Practice differences in inequality

4.1 Does the effect of income on health vary across practices?

We first investigate whether the level of health and the relationship between health and income differs across practices by including practice dummy variables in the regressions for individual health (Table 3). We allow for simultaneity and selection by the method described in section 3. Consequently we use predicted income in the health equation and our sample size is 3477: all individuals who replied to all questions or all but the income questions.

The first, baseline, model is reported in Table 3 and has no practice effects. Model 2 has practice dummies as intercepts and model 3 has both practice dummies as intercepts and interacted with income to allow for the effect of income on individual health to vary across practices. The effects of the individual characteristics other than income are similar across the three models and similar to those in Table 2 which has a smaller sample size and incorporates practice characteristics rather than practice dummies.

The practice fixed effects and income interactions make a significant contribution to explanatory power of the model. The addition of practice main effects (model 2) increases the adjusted R-squared by 7% from 0.138 to 0.148 and including interactions (model 3) increases the proportion of explained variation in health by a further 1% from 0.148 to 0.15. A Wald test rejects the null hypothesis that practice dummies and interactions in model 3 have no significant effect compared with the restricted baseline model 1 ($F(113, 3342) = 1.36$, $Prob > F = 0.0080$). Moreover we also reject the null that model 3 is not significantly different from model 2 (practice main effects only) at the 10% significance level ($F(59, 3342) = 1.25$, $Prob > F = 0.0971$).

Figure 1 shows the distribution of the effects of income on health across the 60 practices. Most practices have positive income effects. The coefficient of variation for the income slopes is 0.518 and ratio of 90^{th} to 10^{th} percentiles of 3.38. The regressions in Table 3 thus provide some evidence that the effect of income on health varies across practices. However, they do not explain why there are differences in the income-health relationship across practices, and they may understate the effect of practices on the relationship if it is affected by a variety of practice characteristics.

4.2 Practice characteristics and income related inequality

Two step procedure

We pursue two different approaches to estimating the association between practice characteristics and the effect of income on practices. First we regress the practice income coefficients from model 3 in Table 3 on practice characteristics. Second we interact practice characteristics with income in a regression for individual health.

Table 4 reports the results from the first procedure in which the estimated coefficients of within practice income effects were regressed on the set of practice characteristics from the QUASAR, GPPAS and GMS statistics. Column 1 is an unweighted regression while column 2 weights the regression by the number of observations in the practice.

Practice characteristics dropped from the regressions after they were found to have no significant association or not to contribute to explaining the variation in income slopes included training status, proportion of GPs performing maternity clinics, the rate of antibiotic prescribing, the score for the treatment of asthma and facilities for the disabled access score and the health authority of the practice.

Some practice characteristics are clearly associated with the income slopes, though the explained variation is not high (23% for the unweighted regression and 13% for the weighted regression). In practices with a higher proportion of female GPs and a lower patient to GP ratio income has a smaller effect on health. The elasticity of the income slope with respect to female GPs is quite small (-9%) whereas the elasticity with respect to the list size per GP is 56%.

Practices with a larger proportion of patients classified as highly deprived had significantly lower income effects, though the elasticity is small (-1%). The characteristic with the largest elasticity (-96%) is the proportion of GPs with out of hours commitments.

One step procedure

Table 5 reports regressions of individual health on individual characteristics, income and the interaction of income with practice characteristics (both centred around their means). Again we have allowed for selection and simultaneity. We also estimated a model with fixed practice effects, individual characteristics, income and interactions of income and practice characteristics to see if the interactions of practice characteristics with income were confounded by unobserved practice effects. The estimated coefficients were very similar.

The results in Table 5 are similar to those in Table 4: income has a greater effect on health in practices with higher rates of antibiotic prescribing and practices with larger list size per GP. The effect of income is smaller in practices who had a greater proportion of GPs providing out of hours care and practices with a greater proportion of patients classified as highly deprived.

5. Discussion

5.1 Simultaneity bias and the measurement of income related inequality

We found that allowing for the simultaneous determination of income and health leads to a doubling of the estimated effect of income on health. This is in line with Ettner who also found that using instrumented income in the health equation led to large increases in the estimated effect of income [12]. Although the simultaneous nature of the income health relationship is well known [13] its implications for the measurement of income related inequality do not appear to have been addressed in the literature. The usual summary measure of income related inequality is the concentration index of health against income [3] which is the product of the Gini coefficient for income and the elasticity of health with respect to income:

$$C_{hy} = \frac{b_{hy}\bar{y}}{\bar{h}} C_{yy} \tag{1}$$

where b_{hy} is the coefficient from the bivariate regression of health on income and C_{yy} is the Gini coefficient for the income distribution (the concentration index of income on income). It is conventional, and of more immediate relevance to policy to decompose the health concentration index to show the direct contribution of income to income related inequality and its indirect contribution via its correlation with other factors affecting health. After estimating a multiple regression of health on income (or a transform of income) and other variables affecting health, the concentration index can be written

$$C_{hy} = \frac{b_{hy\cdot}\bar{y}}{\bar{h}} C_{yy} + \sum_i \frac{b_{hx_i\cdot}\bar{x}_i}{\bar{h}} C_{x_i y} + 2\frac{\mathrm{Cov}(e, F(y))}{\bar{h}} \tag{2}$$

where $b_{hx_i\cdot}$ is the coefficient on x_i from the multiple regression of h on income and other explanatory variables, and $C_{x_i y}$ is the concentration index of x_i against income [14]. (The last term is the generalised concentration index of the regression residual against income and which has a probability limit of zero [1].)

The decomposition provides useful information for policy since it shows how overall inequality would be affected by policies to alter the direct effects of income and the other variables on health, by policies to change the overall distribution of income, and by policies to change the relationship between the other variables such as age or ethnicity and income. However, the multiple regression underlying (2) is typically estimated without allowing for simultaneity between income and health. Our results show that allowing for simultaneity made relatively little difference to the estimated effects of the non-income variables but the income coefficient doubled. Hence the relative importance of the direct effect of income on health is seriously understated if no account is taken of simultaneity, and policy initiatives may be misdirected as a result. Our results suggest that future attempts to decompose income related inequality in health need to take account of simultaneity.

5.2 Practice characteristics and income related inequality

We found that there were significant differences across practices in the effect of income on health and that some practice characteristics had a significant impact on the size of the effect of income on health. It is possible to provide an intuitively appealing story about how some of practice characteristic could affect the relationship between health and income. For example, the rate of antibiotic prescribing can be interpreted as a negative index of prescribing quality [15] and hence of the quality of the practice. The fact that in practices with higher rates of antibiotic prescribing the effect of income on health is greater could therefore be an indication that better quality practices are better able to provide services so that lower income patients are better able to take advantage of them and hence to have better health. We also find that in practices with higher list sizes per GP the effect of income on health is greater. This may be because GPs with lower lists have more time and are thereby able to provider better services which benefit their poorer patients disproportionately. The proportion of GPs who took responsibility for the out of hours care of their patients was also negatively associated with the effect of income on health. This may be because greater continuity of care has a greater effect on the health of the poor than the rich. We also find that a higher proportion of female GPs reduces the effect of income on health, though the coefficient is only significant in the regression of practice income slopes on practice characteristics. Again the effect may be because a higher proportion of female GPs leads to better provision of services to female patients who tend to be poorer than male patients.

It is possible that the practice characteristics are picking up unobserved characteristics of the practice population which influence the effect of income on health. The fact that our results were very similar when we included practice fixed effects, which would pick up such unobserved practice population characteristics, and that we also included a measure of the overall deprivation level of practice patients, may suggest that the practice characteristics are having a genuine, rather than a spurious effect.

The work we report here is the first to explore whether practices can alter the extent of income related inequality by altering the effect of income on health. There do appear to be differences in the relationship between income and health across practices but further studies are required to determine the extent to which practice characteristics are responsible.

References

.

1. Gravelle, H. Measuring income related inequality in health and health care: the partial concentration index, direct and indirect standardisation. *Department of Economics and Related Studies Discussion Paper No.* 17. at http://www.york.ac.uk/depts/econ/dp/2001.htm. August 2001

2. Lambert, P. *The Distribution and Redistribution of Income*, 2nd Edition. Manchester: Manchester University Press, 1993.

3. Wagstaff, A. and van Doorslaer, E. Equity in health care finance and delivery. In: Culyer, A. J. and Newhouse, J. (eds.). *Handbook of Health Economics*. Amsterdam: Elsevier, 2000.

.

4. Campbell, S. M., Hann, M., Hacker, J., Burns, C., Oliver, D. and Thapar, A. Identifying predictors of high quality care in English general practice: observational study. *British Medical Journal* 2001; **323:** 1 – 6.

5. Deaton, A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore and London: John Hopkins University Press for the World Bank, 1997.

6. Stata Corporation. *Stata 7 Reference Manual.* College Station, Texas: Stata Press, 2001.

7. Brazier, J., Roberts, J., and Devrill, M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002; **21**: 271– 292.

8. Angrist, J. D. Conditioning on the probability of selection to control selection bias. NBER Technical Working Paper 181, 1995.

9. Angrist, J. D. Conditional independence in sample selection models. *Economics Letters* 1997; **54:**103 – 112.

10. Angrist, J. D. and Krueger, A. B. Instrumental variables and the search for identification: from supply and demand to natural experiments. NBER Working Paper 8456, 2001. (http://www.nber.org.papers/w8456).

11. Smith, R.J. and Blundell, R.W. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. *Econometrica* 1988; **54**: 679-85.

12. Ettner, S.L. New evidence on the relationship between income and health. *Journal of Health Economics* 1996; **15**:67-86.

13. Smith, J.P. Healthy bodies and thick wallets: the dual relation between health and economic status. *Journal of Economic Perspectives* 1999; **13**: 145-166.

14. Wagstaff, A., van Doorslaer, E. and Watanabe, N. On decomposing the causes of

health sector inequalities with an application to malnuitrition inequalities in Vietnam. *mimeo* 2000.

15. Campbell, S. M., Cantrill, J. A. and Roberts, D. Prescribing indicators for UK general practice: Delphi consultation study. *British Medical Journal 2000;* **321**: 1-5.

Acknowledgements

Funding from the Department of Health to NPCRDC is acknowledged. The views expressed are those of the authors and are not necessarily those of the DH. We are grateful for comments from referees, participants in the York Seminars in Health Econometrics, the 3rd International Health Economics Association conference, and the 10th European Workshop on Econometrics and Health Economics.

HErev1h.doc 14/05/02 18:02 5914

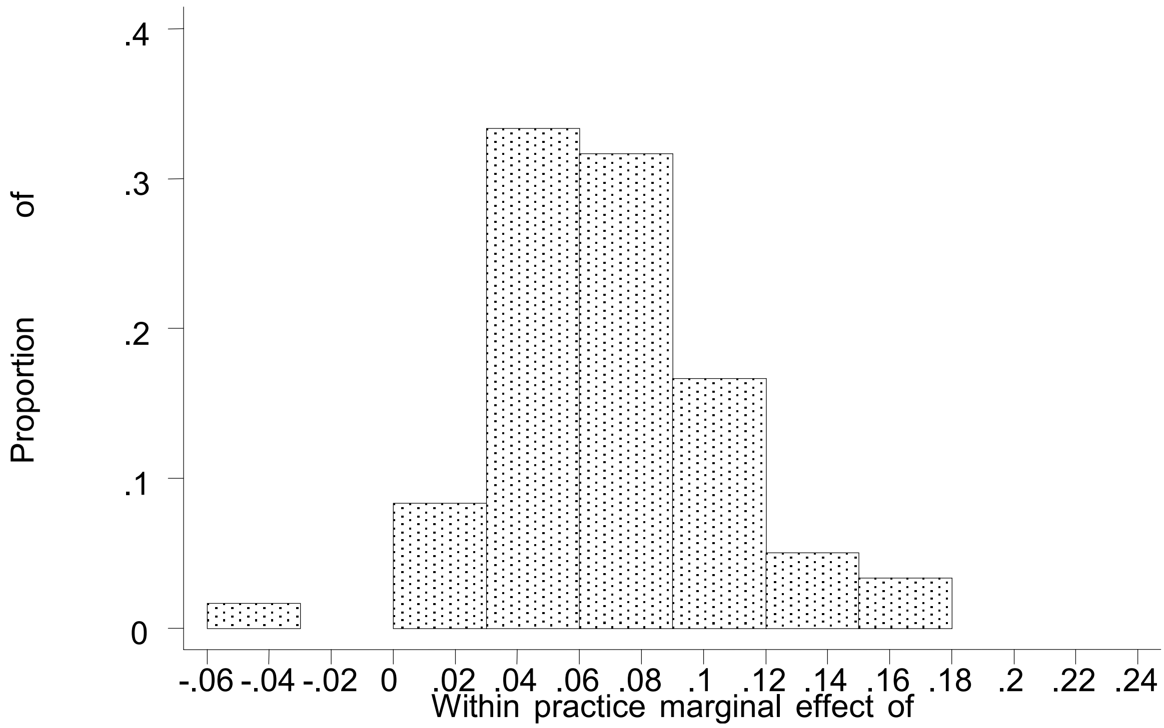

Figure 1. Variation in effect of income on health across practices

Table 1. Variable definitions and descriptive statistics

Variable	Mean	Std. Dev.	Min	Max	Definition
Individual characterisitcs (n = 3477)					
Health	0.807	0.111	0.351	1	SF-6D health state valuation
Income (n = 2340)	22457	15119	999	50000	Household income (midpoint of interval scale)
Female	0.605	0.489	0	1	Female
Age	49.93	16.70	18	99	Age in years
Non_white	0.064	0.246	0	1	Race other than White European
Single	0.123	0.329	0	1	Single
Separate	0.074	0.263	0	1	Separated
Widowed	0.075	0.264	0	1	Widowed
Married (ref)					Married or living together
Non Smoker	0.493	0.500	0	1	Never smoked for more than a year
Children_0 (ref)					No person under 18 in the household
Children_1	0.127	0.333	0	1	One person under 18 in household
Children_2	0.134	0.340	0	1	Two people under 18 in household
Children_3	0.052	0.222	0	1	Three people under 18 in household
Children_4	0.012	0.107	0	1	Four people under 18 in household
Children_5	0.005	0.068	0	1	Five or more people under 18
Practice characteristics (n = 60)					
Training	0.120	0.209	0	1	Proportion of GPs accredited as an approved trainer.
FemaleGP	0.298	0.288	0	1	Proportion of female GPs
Out_hours	0.972	0.135	0	1	Proportion of GPs with out of hours responsibilities
Deprivation	0.014	0.088	0	0.679	Proportion of highly deprived patients
WTEGP	2.820	1.868	1	8	Number of WTE GPs
List_WTEG	2.152	0.525	0.991	3.524	Patients per WTE GP (000s)
Maternity	0.954	0.164	0	1	Proportion of GPs providing maternity medical services
Antibiotic	0.070	0.020	0.041	0.145	Rate of antibiotic prescribing (Items/weighted population)
Disable_score (n = 55)	83.436	16.545	33.33	100	QUASAR disability access index (proportion of 9 access criteria met by practice)
Invmills	0.490	0.190	0.079	1.104	Inverse Mills ratio

Table 2. Allowing for selection bias in income response and for endogeneity

	(1) **Simple model**	(2) **Selection bias**	(3) **Endogeneity**	(4) **Endogeneity and selection**
	Health	Health	Health	Health
Log of Income	0.0295889	0.0298534	0.0636589	0.0720717
	(8.770)**	(9.012)**	(6.528)**	(7.040)**
Female	-0.0152469	-0.0186341	-0.0109911	-0.0295062
	(4.340)**	(2.507)*	(3.134)**	(3.858)**
Age	-0.0139739	-0.0122307	-0.0206181	-0.0119916
	(3.919)**	(2.504)*	(4.824)**	(2.310)*
Age squared	0.0002746	0.0002398	0.0004037	0.0002308
	(3.841)**	(2.452)*	(4.736)**	(2.215)*
Age cubed	-0.0000018	-0.0000016	-0.0000025	-0.0000015
	(4.054)**	(2.742)**	(4.810)**	(2.432)*
Non_white	-0.0029278	-0.0037724	0.007886	0.0052443
	(0.294)	(0.366)	(0.808)	(0.509)
Single	-0.0069756	-0.0062714	0.0099249	0.0173757
	(0.897)	(0.816)	(1.132)	(1.945)
Separated	-0.0224749	-0.0191956	-0.0017572	0.0211982
	(2.674)**	(1.88)	(0.174)	(1.543)
Widowed	0.008029	0.0094687	0.0261465	0.0380514
	(0.742)	(0.854)	(1.957)	(2.598)*
Non_smoker	0.0173634	0.0173662	0.0103724	0.0089728
	(3.624)**	(3.620)**	(1.943)	(1.685)
Children_1	0.0114183	0.0130958	0.0118067	0.0214813
	(1.685)	(1.713)	(1.758)	(2.631)*
Children_2	0.0021064	0.0028859	0.0058549	0.0110735
	(0.253)	(0.339)	(0.692)	(1.249)
Children_3	0.0209605	0.021394	0.0242968	0.0274525
	(2.429)*	(2.493)*	(2.556)*	(2.842)**
Children_4	-0.0030759	-0.0004101	0.0060152	0.0231067
	(0.14)	(0.019)	(0.271)	(1.046)
Children_5	-0.1052266	-0.106628	-0.0941978	-0.0999814
	(3.247)**	(3.269)**	(2.610)*	(2.697)**
Training	0.0180768	0.0197716	0.018694	0.0285145
	(2.788)**	(2.975)**	(2.294)*	(3.287)**
WTEGP	-0.0066562	-0.006319	-0.0094626	-0.0081015
	(2.835)**	(2.563)*	(3.193)**	(2.554)*
WTEGP2	0.0007681	0.0007319	0.0008808	0.0006963
	(3.045)**	(2.720)**	(2.655)*	-1.931
Deprivation	0.0384453	0.038866	0.0466958	0.0507729
	(1.874)	(1.891)	(2.030)*	(2.118)*
Maternity	-0.0387142	-0.0386092	-0.0315194	-0.0294614
	(3.782)**	(3.764)**	(3.568)**	(3.390)**
Antibiotic	-0.3521755	-0.3585348	-0.2524479	-0.2686321
	(2.668)*	(2.695)**	(1.543)	(1.58)
Disability_score	0.0002472	0.0002604	0.0002874	0.000371
	(1.846)	(1.874)	(2.104)*	(2.629)*
Invmills		0.0200147		0.1144969
		(0.505)		(2.625)*
Constant	0.8426313	0.83212	0.8350523	0.7733864
	(36.990)**	(25.610)**	(35.391)**	(22.594)**
Observations	2340	2340	2340	2340

Adjusted R-squared	0.152	0.151	0.104	0.081

Robust t statistics in parentheses

* significant at 5%; ** significant at 1%

Health authority effects fixed effects included in regression were significant at 5% level.

Estimates of the elasticity of health with respect to income (centred on variable means)

	Simple model	**Selection bias**	**Endogeneity**	**Endogeneity and selection**
Income elasticity	0.0365	0.0368	0.0785	0.0889

Table 3. Testing for differences in income related inequality across practices

	(1)	(2)	(3)
	No Practice Effects	Main Effects	Main Effects and Interactions
	Health	Health	Health
Log_Income	0.0739133	0.0708925	.068884¶
	[8.547]**	[7.917]**	
Female	-0.0318614	-0.0261385	-0.0246434
	[4.417]**	[3.663]**	[3.343]**
Age	-0.0145916	-0.0169654	-0.0169120
	[3.728]**	[4.765]**	[4.740]**
Age squared	0.0002819	0.0003273	0.0003237
	[3.561]**	[4.563]**	[4.503]**
Age cubed	-0.0000018	-0.0000021	-0.0000020
	[3.801]**	[4.792]**	[4.707]**
Non_white	-0.0015228	0.0109018	0.0128218
	[0.132]	[1.155]	[1.363]
Single	0.0182426	0.0134115	0.0124706
	[2.262]*	[1.756]	[1.600]
Separated	0.0227783	0.0168557	0.0155474
	[1.954]	[1.432]	[1.303]
Widowed	0.0325407	0.0284301	0.0250712
	[2.892]**	[2.860]**	[2.477]*
Non_smoker	0.0087357	0.0111775	0.0111939
	[1.985]	[2.985]**	[2.944]**
Children_1	0.0181072	0.0127828	0.0111830
	[2.756]**	[2.020]*	[1.744]
Children_2	0.0205358	0.0162091	0.0151285
	[3.030]**	[2.568]*	[2.394]*
Children_3	0.0276268	0.0275980	0.0284421
	[3.740]**	[3.552]**	[3.619]**
Children_4	0.0230220	0.0208574	0.0212125
	[1.243]	[1.096]	[1.094]
Children_5	-0.1274130	-0.0747751	-0.0612368
	[3.149]**	[1.987]*	[1.604]
invmills	0.1025445	0.0704358	0.0657841
	[2.905]**	[1.999]*	[1.783]
Practice main effects		[F(59, 3401) =1.25, Prob > F = 0.0925]	[F(59, 3342) = 1.20 Prob > F = 0.1419]
Practice and income interactions			[F(59, 3342) = 1.25, Prob > F = 0.0971)]
All practice effects			[F(113, 3342) = 1.36, Prob > F = 0.0080)**]
Observations	3477	3477	3477
Adjusted R-squared	0.138	0.148	0.15

Robust t statistics in brackets. * significant at 5%; ** significant at 1%.
¶Mean effect across 60 practices

Table 4. Explaining differences in income related inequality

	(1)		(2)	
	Income slope	Elasticity	Income slope[1]	Elasticity
FemaleGP	-0.0368	-0.15925	-0.02274	-0.09538
	[2.288]*		[2.089]*	
Out_hours	-0.0857	-1.20918	-0.068067	-0.96381
	[4.333]**		[2.015]*	
Deprivation	-0.1009	-0.02098	-0.108091	-0.00990
	[7.665]**		[5.217]**	
WTEGP	0.011906	0.487399	0.010999	0.46667
	[1.541]		[1.801]	
WTEGP2	-0.00123	-0.20281	-0.001186	-0.20809
	[1.420]		[1.685]	
List_WTEG	0.017922	0.559965	0.015026	0.45920
	[2.544]*		[2.557]*	
Antibiotic			0.2815479	0.27281
			[1.602]	
Constant	0.068884		0.0700203	
	[17.016]**		[20.533]**	
Observations	60		60	
Adjusted R-squared	0.229		0.133	

Robust t statistics in brackets
* significant at 5%; ** significant at 1%

[1] Estimates using weights proportional to the number of observations within each practice

Table 5. Regression of individual health on individual characteristics, level of practice characteristics, and interaction of practice characteristics with income.

	Clustering within practices
Log_Income	0.0723717
	[8.343]**
Female	-0.0293167
	[4.000]**
Age	-0.0147978
	[3.950]**
Age squared	0.0002851
	[3.759]**
Age cubed	-0.0000018
	[3.999]**
Non_white	0.0020474
	[0.172]
Single	0.0169789
	[2.131]*
Separated	0.0220592
	[1.899]
Widowed	0.0305418
	[2.641]*
Non_smoker	0.0093924
	[2.228]*
Children_1	0.015954
	[2.458]*
Children_2	0.0192911
	[2.894]**
Children_3	0.0269051
	[3.733]**
Children_4	0.0229908
	[1.253]
Children_5	-0.1261178
	[3.044]**
Training	0.0277354
	[4.322]**
WTEGP	-0.0061709
	[2.035]*
WTEGP2	0.0005734
	[1.676]
List_WTEG	0.0000006
	[0.144]
Deprivation	0.0325705
	[1.683]
Out_hours	-0.0206748
	[1.367]
FemaleGP	-0.0124626
	[1.605]
Maternity	-0.015918
	[1.036]
Antibiotic	-0.3946417
	[3.155]**
Inc_Training	0.0168768
	[1.609]

Inc_WTEGP	0.0089332
	[1.764]
Inc_WTEGP2	-0.0008932
	[1.557]
Inc_List_WTEG	0.0105978
	[2.132]*
Inc_Deprivation	-0.0867176
	[4.933]**
Inc_Out_hours	-0.0670579
	[2.122]*
Inc_FemaleGP	-0.0072745
	[0.776]
Inc_Maternity	-0.0097406
	[0.622]
Inc_Antibiotic	0.3484279
	[2.688]**
Invmills	0.091246
	[2.566]*
Constant	0.7566318
	[46.191]**
Observations	3477
Adjusted R-squared	0.144

Robust t statistics in brackets
* significant at 5%; ** significant at 1%