Do GPs improve health? Cross-sectional analysis of the association between individual-level health status and GP supply

Stephen Morris[a] **Hugh Gravelle**[b] **Matthew Sutton**[c]

[a]*Tanaka Business School, Imperial College London; NPCRDC, University of Manchester.*
[b]*NPCRDC, Centre for Health Economics, University of York; hg8@york.ac.uk*
[c]*General Practice and Primary Care, Community-Based Sciences, University of Glasgow; ISDScotland*

Abstract. We use data from four rounds (1997-2000) of the Health Survey for England to examine whether the supply of GPs has an effect on the self assessed health of individuals. We regress individual health status on GP supply plus covariates, using three measures of GP supply and five measures of health. When no allowance is made for endogeneity, GP supply has a small, statistically insignificant impact on health status which is negative in 6 of the 15 models. Using instrumental variables methods to allow for the endogeneity of GP supply, we find that GP supply has a much larger positive effect which is statistically significant in 12 of the 15 models. The elasticity of health with respect to GP supply ranges from 0.19 to 0.44 depending on the health and GP supply measures used. The incremental cost per QALY gained following a 10% increase in GP supply ranges from £230 to £531.

Key words: Health. GPs. Primary care. Instrumental variables.

1 Introduction

There is a growing debate about whether health care improves population health (Davey Smith et al, 2000). Despite early, influential studies indicating little role (McKeown, 1976), it has been argued more recently that medical advances that can be applied on a substantial scale have led to a more important role for health care in improving health (Bunker 1995; Watt, 2003; Cutler and McLellan 2001).

While there is a growing evidence base for many technologies, there have been relatively few studies of the aggregate effect of health care on health. Some studies have adopted 'inventory approaches' and argue that applying treatment effects to aggregate measures of activity shows a considerable health effect (Capewell et al, 1999; Tunstall-Pedoe et al, 2000). Cross-national studies of the association between health care expenditure and health, however, find little evidence of a positive effect. The Wanless Report and the scenarios for future health and health care expenditure make understanding the role of health care in health improvement all the more important.

A consistent focus of recent health policy has been to strengthen the role of primary care. It has been argued (Macinko and Starfield, 2003) that systems with better primary care services have better health. In this paper we investigate whether a higher supply of family doctors is associated with better population health using a multi-level (area and individual) data set for England.

A number of recent studies have investigated the relationship between the delivery of primary care and health indicators. The majority of studies have been conducted at the ecological level, relating aggregate area level measures of primary care delivery to population health outcomes.

National level. Starfield and Shi (2002) used cross-sectional data on 13 countries and found that a measure of the strength of primary care infrastructure had negative bivariate correlations with health care costs and positive bivariate correlations with health indicators. Macinko, Starfield and Shi (20003) constructed a panel of 18 OECD countries over the period 1970 to 1998. They used fixed effect panel data multiple regressions and found that the strength of a country's primary care system was negatively associated with mortality (both all cause and a number of specific causes) after controlling for per capita GDP, physicians per capita, proportion of elderly, ambulatory care visits, tobacco and alcohol consumption.

Area level. Shi, Starfield, Kennedy and Kawachi (1999) used path analysis (structural equation modelling) at US state level to examine the relationships between primary care doctors per capita, income inequality, speciality care, smoking, and mortality based health indicators. Primary care provision had beneficial effects on all the health indicators but areas with greater income inequality had lower provision. Shi and Starfield (2001) used cross sectional data from 273 US Metropolitan Areas in 1990 and found that, after allowing for per capita income, education, unemployment, urban population proportion, percentage black and the Gini coefficient, the primary care physician to population ratio was significantly negatively associated with black total mortality and insignificantly negatively associated with white mortality. Jarman et al (1999) used data on 183 hospitals and found that inpatient mortality rates were lower in hospitals with lower proportions of cases admitted as emergencies, higher ratios of consultants to beds and higher numbers of GPs per capita. Gulliford (2002) used data from 99 English Health Authorities (HAs) in 1999 and found that HAs with more GPs per capita had lower all cause and specific mortality, lower hospital admissions and lower conceptions under 18, after allowing for Townsend deprivation scores, proportion of ethnic minorities and lower social class proportions.

Multi-level. Shi and Starfield (2000) found that individuals were more likely to report good health if they lived in states with more primary care doctors per capita (and with higher income inequality), after controlling for gender, age, ethnicity, employment, wage, whether the individual was below the poverty line, heath insurance, SF12 physical health score and whether the individual had ever smoked. The data were from a 1996 sample of 58,000 clustered in 60 communities. Shi, Starfield, Politzer and Regan (2002) used the same data source but in addition made use of responses to questions about accessibility of primary care, interpersonal care and continuity of care. The results were similar to Shi and Starfield (2000) in that better primary care (and greater income inequality) were found to be associated with better physical and mental health after controlling for a wide range of covariates.

Dusheiko, Gravelle and Campbell (2003) examined the relationship between individual level health and practice characteristics for a sample of 2500 individuals clustered in 60 practices in 6 HAs in 1998. They found female patients in practices had better health the greater the proportion of female GPs but that the proportion of GPs with non-UK qualifications had no association with the health of non-whites. Practices with characteristics indicating higher quality had healthier patients, but there was no direct effect of list size per GP on health.

This account of the literature suggests that there have been only three UK studies of the relationship between primary care provision and population health. One was an ecological study using relatively few observations, a very simple measure of general practice provision, and a small number of covariates. The second was also ecological and concentrated only on inpatient mortality. The remaining study was multi-level but was based on a relatively small sample with a low response rate and considerable non-response for income. The other two multi-level studies we have found are from the US and make use of the same database.

The major problem in investigating the relationship between individual health and the supply of health care is endogeneity: we need to allow for the possibility that supply is partly determined by health. It is plausible that GPs like to live and work in "nice" areas and "nice" areas have unobserved characteristics which lead them to have healthier populations. This could lead to a positive estimated effect of GPs on health even if GP supply has no true effect. GPs' location decisions are also regulated and influenced by a national payment system. Until the recent reforms, the distribution of the GP workforce was regulated by the Medical Practices Committee, whose remit was to ensure 'adequate' provision in each area. The evidence suggests that this regulation was somewhat unsuccessful (Gravelle and Sutton, 1998, 1999), probably because it was generally reactive rather than proactive. The payment system (Standard Fees and Allowances) provides financial incentives for practicing in areas with high workload through deprivation payments and some fee-for-service elements (such as night visits). Such considerations would lead us to expect a negatively biased effect of GP supply on health, as various policy mechanisms try to encourage GPs to locate in poor health areas or prevent them from preferentially locating in good health areas.

The most robust method of estimating the true effect of GP supply on health is to use panel data in which there are large enough changes in GP supply over time. But GP supply changes relatively slowly so that a long panel would be required and the longer the panel the less confident one could be that any unobserved area characteristics affecting health and GP supply are constant over time and can be washed out by observing areas over time. In this paper we follow the alternative strategy of attempting to remove endogeneity bias by finding instrumental variables which are correlated with GP supply but uncorrelated with unobservable factors affecting individual health.

2 GP supply and health

We start by setting out a simple model of health to guide our estimation of relationship between health and GP supply. The general self assessed health (h) of individual i in area a is determined by their socio-economic characteristics, x_{ia} and their use of health care u_{ia}

$$h_{ia} = f^{h}(x_{ia}, u_{ia}, e^{h}_{ia}) \quad (1)$$

where e^h is an error term. But their consumption of health care depends on their health status, their socio-economic characteristics, the accessibility of health care and an error term reflecting all the unobservable factors affecting use:

$$u_{ia} = f^{u}(h_{ia}, x_{ia}, g_{ia}, c_{ia}, e^{u}_{ia}) \quad (2)$$

g_{ia} is a measure of the accessibility of GP services and c_{ia} is a vector of measures of the accessibility of other types of health care in the area, such as distance to hospitals etc. Any attempt to measure the effect of health care use on health by estimation of (1) will be vitiated by simultaneity: the estimated effect of health care use will be biased because it will pick up both the effect of health care on health and the effect of health on use (sicker people have more use). Moreover, in the surveys which yield information on self assessed health, the utilisation variables are usually also self reported and may be only for the period immediately preceding the survey. We use the Health Survey for England (HSE) because it has a very rich set of health measures. However, the HSE asks about the use of GP services in the previous fortnight and does not contain information on the quality of services provided.

We investigate the relationship between general self assessed health and the supply of one type of health care (GP supply) by considering the reduced form health model obtained by substituting (2) into (1): $h_{ia} = f(x_{ia}, g_{ia}, c_{ia}, e_{ia})$. However, we do not have sufficient information to measure the accessibility of services for each individual (g_{ia}, c_{ia}). This would require, for example, data on the distances from individuals to providers, their car ownership, their costs of waiting etc. Instead we measure accessibility at area level by using the numbers of GPs in an area, the average distance from area centroids to providers, the average waiting times at providers etc. To fix ideas, suppose that the reduced form heath equation is linear

$$h_{ia} = \eta_0 + \eta_1 x_{ia} + \eta_2 g_a + \eta_3 c_a + e_{ia} \qquad (3)$$

where e_{ia} is an error term reflecting all the unobserved variables affecting health and the supply variables g_a, c_a are measured at area level.

The reduced form health equation is of interest in its own right because the coefficient on the GP supply variable gives the effect of increasing supply on general self assessed health. Thus one may calculate whether increasing supply is a more or less cost-effective method of attempting to improve health than other policy interventions.

If e is uncorrelated with g then OLS will produce an unbiased estimate of the impact of GP supply on general self assessed health. But GPs may choose to locate in areas with many amenities and such areas may have populations which are more likely to be healthy. Or health authorities with sicker populations may attempt to recruit more GPs. Thus GP supply may be correlated with area health and hence with individual health, leading to biased estimates of the effect of GP supply variable on health. We have a very rich set of individual level morbidity and socio-economic characteristics. This will reduce the extent of any bias but it will not eliminate it.

One method of dealing with such bias is to find instrumental variables (IVs) which do not enter the health equation directly, are not correlated with the error term in the health equation and can predict GP supply. It is possible (see section 3) to test whether GP supply is exogenous. If it is not then we can look for instrumental variables. It is also possible to test whether putative instrumental variables are able to predict GP supply but not whether they are uncorrelated with the errors in the health equation which are by definition unobservable. Instead one must rely on other arguments to establish the plausibility of the instruments.

Capitation payments are a major component of GP income and so, ceteris paribus we expect to find more GPs in areas where the population generate higher capitation payments. We capture such higher payments per head of population indirectly by the proportion of elderly patients (since capitation payments are higher for patients aged over 75) and directly by calculating the average capitation payments implied by the area age structure. We argue that these satisfy the second requirement of valid instruments of being uncorrelated with unobservable factors affecting individual health. Although older patients are less healthy than younger patients we include an age variable in the health equation (3), hence age is not a component of the error term and does not give rise to a correlation between GP supply and the error term. It is then difficult to think of any reason why the health of an individual patient, given their age (and all the other socio-economic and morbidity variables included in (3)), should be correlated with the age structure of the area.

Since we are concerned with estimating the effect of GP supply on health we do not worry about the possibility that the variables measuring access to other types of health care are correlated with the error terms in the health equation. Any correlation between the error term in the health equation and our measures of access to other types of care will bias the effects on these coefficients, not the coefficient on the instrumented GP supply variable which is by assumption uncorrelated with the error in the health equation.

3 Data

We combine data from a number of sources to generate alternative measures of the health of individual i in area a (h_{ia}) and of the supply of GPs in area a (g_a).

3.1 Sources

The analysis is based on individual-level data from the Health Survey for England (HSE). The HSE is a nationally representative survey of individuals aged two years and over living in England. A new sample is drawn each year. We use four rounds of the HSE from 1997 to 2000 giving a total of around 65,000 observations depending on which health measures we use. The HSE has a rich set of individual socio-economic and health variables.

We construct area level GP supply variables using the AREA project dataset and the National Primary Care Research and Development Centre (NPCRDC) general medical services dataset. The AREA dataset was assembled for a project which examined the determinants of the utilisation of hospital and community health services and general practice prescribing for a review of the resource allocation formulae for England (Sutton et al., 2002; Gravelle et al., 2003).

3.2 Health

We used a number of individual-level health variables from the HSE to construct measure of health:

- Self-assessed general health, in five categories: “very bad”, “bad”, “fair”, “good”, and “very good”;
- Acute ill health, measured by the number of days in the last two weeks affected by acute illness (0, 1-3, 4-6, 7-13, 14);
- Specific longstanding illnesses (15 binary variables by broad disease code)
- Limiting longstanding illness (LLI), a binary variable defined by having at least one longstanding illness which is limiting;
- Number of longstanding illness (0 or 1, 2, 3, 4 or more);
- GHQ-12 scores measured as ordered categories (0-12);
- Economic inactivity due to ill health (intending to look for work but prevented by temporary sickness or injury, permanently unable to work because of long term sickness or injury, retired – we control for age and interpret this last category as those taking early retirement due to ill health).

We constructed five measures of health status whose form determines the statistical model used to estimate the relationship between health and GP supply.

Self-assessed general health, h^1

We estimate an ordered probit model for the responses $h^1_{ia} = 1,2,3,4,5$ to the self assessed general health question where

$$h^1_{ia} = j \text{ if } \sigma_{j-1} < h^{1*}_{ia} \le \sigma_j, \quad \text{for } j = 1,2,...,5, \quad \sigma_0 = -\infty, \sigma_5 = +\infty \qquad (4)$$

where the latent variable h^*_{ia} is determined by GP supply and covariates as

$$h^*_{ia} = \varsigma_1 x_{ia} + \varsigma_2 g_a + \varsigma_3 c_a + \kappa^1_{ia}, \quad \kappa^1_{ia} \sim N(0,1) \qquad (5)$$

Discrete attributed EQ-5D score h^2

We used the 1996 HSE sample which has information on EQ-5D scores to compute the mean EQ-5D score for the individuals in the *j*-th self assessed general health categories and assigned it as the health score for individuals in category *j* in the 1997 to 2000 rounds of the HSE. We treat h^2 as a cardinal variable and estimate the health model by least squares regression.

Continuous attributed EQ-5D score h^3

Although h^2 is a cardinal measure it takes on only five values. We use a method for extracting more information from the data which yields an EQ-5D score for each individual which is not constrained to five values. The method was proposed and validated on good quality Canadian health data by van Doorslaer and Jones (2003, p.65) and subsequently applied to UK data by Sutton (2002) and Gravelle and Sutton (2003). The 1996 HSE data on EQ-5D scores and self-assessed general health are used to set the cut points for the self-assessed health categories (the σ's in (4)). The cumulative percentages of respondents in the 1996 HSE reporting "very bad" to "good" health are 1.16%, 5.38%, 23.40%, and 64.66%. Corresponding values from the empirical distribution of EQ-5D scores are -0.016, 0.364, 0.743 and 1.000. If it is assumed that the mapping of latent self assessed general health to EQ-5D scores is stable over time, then the unobserved EQ-5D scores (h^3) for respondents in the 1997 to 2000 HSEs and their reported self assessed general health responses are related by

$$h^1_{ia} = j \text{ if } \sigma_{j-1} < h^3_{ia} \le \sigma_j, \text{ for } j = 1,2,...,5 \qquad (6)$$

where $\sigma_0 = -\infty, \sigma_5 = \infty$, $\sigma_1 = -0.016$, $\sigma_2 = 0.364$, $\sigma_3 = 0.743$, and $\sigma_4 = 1.00$. With the EQ-5D score determined by

$$h^3_{ia} = \phi_1 x_{ia} + \phi_2 g_a + \phi_3 c_a + \kappa^3_{ia}, \quad \kappa^3_{ia} \sim N(0,1) \qquad (7)$$

we can use interval regression to estimate the effects of GP supply and the covariates on health measured by the EQ-5D score.

Predicted linear index from an ordered probit regression of self-assessed general health against health variables, h^4

This method is similar to the ordered probit model used in h^1 except that we assume that the latent variable h^{1*}_{ia} depends directly only on the morbidity variables (and age and gender) m_{ia} in the HSE:

$$h^{1*}_{ia} = \iota m_{ia} + \kappa^4_{ia}, \quad \kappa^4_{ia} \sim N(0,1) \qquad (8)$$

Thus it is assumed that the other variables, such as GP supply, income etc affect self assessed general health only indirectly through their effect on the other health variable in (8). We can regard h^{1*} as a convenient summary of all the health variables.

From an ordered probit regression of self assessed general health on the other health variables, age, and gender we obtain the predictions of the linear index as $\hat{h}^{1*}_{ia} \mid \hat{\iota} m_{ia}$. We follow Van Doorslaer and Jones (2003) in rescaling the measure to lie within the [0, 1] interval

$$h^4_{ia} \mid \left(\hat{\iota}\, m_{ia}\ 4 \,/\, \hat{\iota}\, m\right)^{\min} \Big\}\Big(/\, \hat{\iota}\, m\Big)^{\max} 4 \,/\, \hat{\iota}\, m\Big)^{\min} \Big\}^{41} \qquad (9)$$

where $(\hat{\iota}\, m)^{\max}$ and $(\hat{\iota}\, m)^{\min}$ are the largest and smallest values of the sample individual linear predictions, respectively. Other rescaling options are possible (including no rescaling). The scale merely affects the magnitude of estimated coefficients, not their significance. We then use this as the health measure in a regression of health on GP supply, income etc. The ordered probit model specified in (8) is in Appendix 1. The health indicators are generally statistically significant and have the expected sign.

Predicted EQ-5D score from an interval regression of self-assessed general health against health variables, h^5

This method is a combination of methods used in h^3 and h^4. We first estimate an interval regression model similar to (6) and (7) where the unobserved latent EQ-5D health variable h^{3*} now depends directly on health variables and demographic variables m

$$h^3_{ia} \mid \varpi m_{ia}\ 2\ \kappa^5_{ia}, \quad \kappa^5_{ia} \sim N(0,1) \qquad (10)$$

h^5 is obtained from the individual predictions of the linear index of this model

$$h^5_{ia} \mid \hat{\varpi} m_{ia} \qquad (11)$$

and can be used as the health measure in least square regression on GP supply, socio-economic variables, and secondary care access variables. The interval regression model specified in (10) is in Appendix 1.

Table 1 give summary statistics for the five health measures and Table 3 summarises their relative advantages and disadvantages.

3.3 GP supply

GP supply as weighted average of practice list size: g^1

We constructed a series of area level GP supply variables using the Attribution Data Set (ADS) for 2000 from the AREA data set. The ADS gives the number of patients on a practice list who were resident in each ward a in England in April 2000. Denote this $N_{ap}(2000)$ We have data on the total list size ($N_p(t)$) for each practice p in each year t and the total whole time equivalent (WTE) number of GPs in each practice $G_p(t)$. Let $S_p(t) \mid G_p(t) / N_p(t)$ denote the number of GPs per patient in practice p in year t. Our first measure of GP supply in ward a in year t is

$$g^1_a(t) \mid \sum_p S_p(t) \frac{N_{ap}(2000)}{\sum_p N_{ap}(2000)} \qquad (12)$$

GP supply as weighted average of ward list sizes: g^2

To construct an alternative ward level measure we focus on the ward within which a practice is located. From the ADS we get $N_{alk}(2000)$, the number of people in 2000 who lived in ward *a* who were registered with practice *k* located in ward ? . We then calculate the total number of people in ward *a* who were registered with a practice located in ward ? as $N_{al}(2000) = \sum_k N_{alk}(2000)$. From the NPCRDC database we have the number of people registered with the *k*th practice located in ward ? in year *t*: $N_{?k}(t)$. Let $G_{?k}(t)$ be the number of GPs in the *k*th practice in ward ?, so that the average list size per GP in practices located in ward ? is

$$S_?(t) = \sum_k G_{?k}(t) / \sum_k N_{?k}(t) \qquad (13).$$

We calculate the second ward level supply variable as

$$g_a^2(t) = \sum_? S_?(t) \frac{N_{a?}(2000)}{\sum_{?\Re} N_{a?\Re}(2000)} \qquad (14)$$

GP supply at Local Authorities level: g^3

We compute a GP supply measure at a higher level of aggregation: the local authority of the patient. Each practice *p* is located within a local authority *b*, of which there are 354. We use the NPCRDC database to find for the *p*'th practice in local authority *b* the number of patients in each year (N_{bp}(t)), the number of WTE GPs ($G_{bp}(t)$) and the list size: $S_{bp}(t) = G_{bp}(t)/N_{bp}(t)$. Our third GP supply measure is

$$g_b^3(t) = \sum_p S_{bp}(t) \frac{N_{bp}(t)}{\sum_p N_{bp}(t)} \qquad (15)$$

Table 2 has summary statistics for the three GP supply measures and Table 3 lists some of their disadvantages. The local authority level measure has the advantage that it does not require the assumption that the distribution of ward populations across practices is constant over the period 1997 to 2000. However, it suffers from the disadvantage that GP supply varies within local authorities so that it may be too coarse a measure given that individuals will be affected by the provision of GP services close to where they live. Variation within local authorities across wards accounts for 48% and 44% of the total variation in the ward level supply variables g^1 and g^2, respectively.

We estimate the impact of GP supply in the current year on current health. This assumes that (a) there is no lag in the impact of GP supply on health; or (b) there is a lag but the GP supply variables are highly correlated across years.

3.4 Covariates

We include the following individual level covariates in the health equation

- Age and sex, measured as age in years/100, age squared, age cubed, female, female*age, female*age squared and female*age cubed
- Income, measured as equivalised banded total household income where the household income value for each individual is defined as the median value in each band computed assuming a log normal distribution
- Social class of head of household (I professional, II managerial technical, IIIN skilled non-manual, IIIM skilled manual, IV semi-skilled manual, V unskilled manual, other)
- Highest educational qualification attained (degree or equivalent, higher education qualification less than a degree, A level or equivalent, GCSE or O level or

equivalent, CSE or equivalent, other qualifications, no formal educational qualifications)
- Ethnicity (White, Black Caribbean, Black African, other Black ethnic group, Indian, Pakistani, Bangladeshi, Chinese, other non-White ethnic group)
- Rurality, measured using the individual level categorical variable in the HSE (degree of urbanisation = rural, suburban or urban)
- Number of cars owned by household (0, 1, 2, 3 or more)
- Marital status (single, married and living with husband or wife, married and separated from husband or wife, divorced, widowed)
- Household tenure (own outright, buying with a mortgage, part rent part mortgage, rent, live rent free)
- HSE year dummies
- Month of interview (11 dummy variables)
- Dummy variable for proxy respondent (in the HSE responses for children aged 12 years or younger are obtained from the parent or guardian with the child in attendance)

We also include a number of variables from the AREA project database which measure the accessibility of secondary care since use of secondary care is also likely to affect health.

- Hospital distance, measured as the average distance to the five nearest acute care providers from the centroid of the ward of residence
- Hospital beds, measured as the average beds number of beds at acute providers used by individuals in the ward of residence
- Hospital waiting times, measured as the average proportion of outpatients waiting less than six months for treatment at providers used by individuals the ward of residence

We also included dummy variables for the Government Office Region of residence of the individual. This pick up any unobserved regional supply conditions and other factors affecting health. The sample size was maximised by imputing missing items for all the independent variables. Dummy variables are included for each imputed item to indicate item non-response.

4 Estimation Methods

When health status is measured as a continuous variable (as in the case of h^2, h^4 and h^5) we use two stage least squares (2SLS) to allow for the potential bias arising from the endogeneity of area GP supply. We first estimate a reduced form equation for GP supply as

$$g_{ia} = \zeta_0 + \zeta_1 z_{ia} + \zeta_2 x_{ia} + \zeta_3 c_a + e^g_{ia} \qquad (16)$$

where z is a vector of variables that are correlated with g and, we assume, uncorrelated with the unobserved factors affecting individual health. To test for endogeneity we use the Hausman test (Hausman, 1978). We estimate (16), compute the residuals, and add them to the health equation (3) as explanatory variables. Under the null hypothesis that GP supply is exogenous the coefficient on the residuals will be zero. If the coefficient is significantly different from zero the null is rejected and IV methods should be employed (Wu, 1973; Hausman, 1978).

To get an unbiased estimate of the effect of GP supply (on the assumption that the instruments z are not correlated with the error in the health equation) we replace g in the health equation (3) with its predicted value $\hat{g}$ from the reduced form GP supply equation

(16). Note that g^3 is a local authority-level variable and the instruments z are ward-based. Hence ward-level variation is introduced into $\hat{g}^3$.

It is impossible to test empirically whether the instruments are exogenous, but when we have more instruments than we need to identify an equation it is possible to test the conditional validity of the additional instruments with the Sargan test (Sargan, 1958). Let r be the number of endogenous right hand side variables, and s be the number of instruments. If $s > r$ and at least r instruments are exogenous, we can test the null hypothesis that all the instruments are exogenous against the alternative hypothesis that at least one but no more than $s–r$ is endogenous by regressing the residuals from the health equation estimated with the r instruments against x, c, and z. Under the null hypothesis x, c, and z should have no explanatory power and the R^2 of the regression (call this R_u^2) should be zero. The test statistic is nR_u^2 where n is the number of observations. This is distributed as θ^2 with $s–r$ degrees of freedom. A rejection means that the full instrument set is jointly invalid, though individual instruments may still be valid. Failure to reject the null hypothesis means that we can have "some confidence" (Wooldridge, 2002, p.123) in the overall set of instruments used. However, the result is predicated on the assumption that at least r instruments are exogenous.

If the IV models are estimated by OLS the standard errors will be incorrect since they do not take into account the two-step nature of the estimation (Greene, 2000; Wooldridge, 2002). In the results the z-scores for each regressor are reported based on the asymptotic covariance matrix given in Wooldridge (2002, p.95).

Although two of the health measures (h^1, h^3) are not continuous variables, Wooldridge (2002, p. 507) suggests that one can still test for endogeneity using the same approach as for continuous variables. It is also possible to allow for the endogeneity of GP supply using IV methods by first estimating the GP supply equation by OLS and then using the predicted supply ($\hat{g}$) in the second stage regressions even though the models for h^1 and h^3 are estimated as ordered probit and interval regressions. The standard errors from health equation are invalid because they do not take account of two-stage nature of the estimation process (Wooldridge, 2002, p. 475). We therefore compute bootstrap estimates for the standard errors of the coefficients in the second stage regressions. These are the standard deviations of the coefficients from 50 replications of both stages of the estimation based on random draws from the data of n observations with replacement for each replication.

The models are estimated using Stata version 8.2.

5 Results

Summary statistics for the health status measures are shown in Table 1. The proportions of the sample reporting very good, good, fair, bad and very bad self-assessed general health are 37.8%, 40.3%, 16.6%, 4.1% and 1.2%, respectively. For h^2, h^4 and h^5 the mean health status value across the whole sample is 0.85, 0.82 and 0.84, respectively, and the range is within acceptable boundaries (0.21 to 0.94, 0.00 to 1.00, and -0.12 to 0.99, respectively). The mean value within each category of self-assessed general health status is decreasing in worsening health. Note that for h^2 the standard deviation within each category is zero since all observations within the category have the same value.

Table 2 contains individual-level (n=66,484) summary statistics and correlation coefficients for the GP supply measures. For all measures the mean value is similar, with around 0.50 WTE GPs per 1,000 registered persons, or one WTE GP for every 2,000 people. The median value for each measure is the same as the mean, which suggests that the distribution of GP supply is symmetric. The measures of dispersion for g^1 and g^2 suggest these variables are very similar. There is less variation in g^3 than in g^1 or g^2, and given that the measures of central tendency plus the 5th and 95th percentiles and the interquartile range are similar for all measures this suggests that g^3 has fewer extreme values at both ends of the distribution. This is unsurprising since this measure is based on larger geographical areas than g^1 or g^2. Excluding outliers, we find considerable variation in GP supply – the ratio of the 95th to 5th percentile is between 1.3 and 1.4. The bottom panel of Table 2 shows that the measures are positively correlated, and the correlation coefficient is highly statistically significant.

Estimates of the impact of GP supply on individual health which do not allow for the endogeneity of GP supply are presented in Table 4. The table contains the results for each measure of health status regressed against each GP supply measure plus covariates. g^1 and g^2 have negative coefficients in the models with health status measured by h^1, h^2 and h^3 and positive coefficients in those with h^4 and h^5. GP supply measured by g^3 exerts a positive effect on health status for all measures. In none of the models is the impact of GP supply on health statistically significant.

The IV regression results are presented in Table 5. We report the joint significance of the two instruments in the first stage regressions conditional on the covariates using an F test. A rejection ($p < 0.05$) suggests that the instruments satisfy the non-weakness requirement. Staiger and Stock (1997) suggest that ten is the minimum acceptable value of the F statistic computed under the null hypothesis that the instruments are not partially correlated with GP supply once the other exogenous variables have been netted out. On this basis the instruments are non-weak predictors of GP supply conditional on the covariates for all models. For the 2SLS models (h^2, h^4 and h^5) we also compute the Sargan θ^2 statistic to examine the orthogonality of the instruments. Failing to reject the null hypothesis suggests that the full instrument set is jointly valid. The orthogonality of the instruments is rejected in all cases for health measure h^2, although this may reflect general model misspecification (Godfrey and Hutton, 1993) because of the discrete nature of the dependent variable. For h^4 and h^5 we fail to reject the null hypothesis for all three GP supply measures, suggesting that the overidentifying instruments satisfy the orthogonality requirement.

We also report the results of the Hausman exogeneity tests for GP supply and health status. For h^2, h^4 and h^5 we report the F statistic under the null hypothesis that the coefficient on the residuals term is zero. For h^1 and h^3 we report the z score. In both cases a rejection indicates that the residual term is non-zero and hence GP supply should not be treated as endogenous and IV models are required. Exogeneity is rejected for all GP supply measures with health status measured by h^2, h^3, h^4 and h^5. Using the IV ordered probit model for h^1 we fail to reject the exogeneity of GP supply and health.

By contrast with the models estimated without the use of IV methods, GP supply has a positive effect on health status. The effect is statistically significant for h^2, h^3, h^4 and h^5. The models estimated with the local authority supply measure have the largest coefficients on

GP supply for a given health measure. One explanation is that the local authority GP supply variable has a smaller range than the ward level supply variables so that the per unit effect of an increase in supply is greater: the partial slope of the regression line is steeper.

The health variables are measured in the same EQ-5D units in models with h^2, h^3, h^5 so that the coefficients can be compared across models. The model with h^2 has the lowest coefficient but appears to be misspecified. The coefficient in the h^3 model is larger than in the h^5 model, suggesting that most (70%-84%) of the effect on health is in the more 'objective' measures with the remainder being in self-assessed health only. This remainder may be interpreted as measurement error or unmeasured components of morbidity.

Appendix 2 contains the full results for two of the IV models: regressions of health measures h^3 and h^5 against GP supply measure g^1 plus covariates. We report both stages of the estimation. There is a small difference in the sample sizes (the h^3 model excludes observations with missing self-assessed general health). The two instruments are significant predictors of GP supply. GP supply is positively correlated with the proportion of the ward population aged 75 or over. It is negatively correlated with the ward age-related capitation payment, conditional on the proportion of the ward population aged 75 or over. The difference in the sign of the coefficients is possibly due to multi-collinearity; the correlation coefficient between the two variables is 0.62 ($p < 0.0001$). GP supply has a weak positive correlation with individual income. Relative to those with a degree those with no education qualifications have lower GP supply. Relative to Whites, the Black Caribbean, Pakistani and Bangladeshi ethnic groups have greater GP supply. Individuals living in urban or suburban areas have significantly lower GP supply than those in rural areas. GP supply is positively correlated with the distance to secondary care, the number of beds at secondary providers and shorter waiting times.

In the second-stage health equations, health status is positively correlated with income. The coefficient in the h^3 model is larger than in the h^5 model. We also find evidence from both models that health status is positively correlated with higher educational attainment, the number of cars owned, being married, and owning a house outright. There are some notable differences between the two models. Conditional on the other covariates lower social class has an insignificant effect in the h^5 model while for h^3 there is a significant and negative effect. Relative to Whites, the non-White ethnic groups have lower subjective health status, while in terms of the more objective measure h^5 the impact is mainly insignificant. The Chinese group have significantly higher objective health than the White group. Distance to secondary care is negatively correlated with health in the h^3 model, and insignificant in the h^5 model. The number of beds exerts a significant and negative effect on h^5. Generally the coefficients are of a smaller magnitude in the h^5 model.

For the models with h^3 and h^5 the elasticity of health with respect to GP supply ranges from 0.19 to 0.44, depending on the GP supply measure used. For example, in the model with h^5 and g^1 the elasticity is 0.19: an increase in GP supply of 10% will increase the EQ5D score by 1.9%, which is, estimated at the sample means, an increase of (1.019*0.84)-0.84 = 0.016 EQ-5D units.

It is possible to make a crude estimate of the cost per quality adjusted life year (QALY) gained from an increase in GP supply. In the model using the health measure h^5 and GP supply measure g^1, the absolute change in health estimated at the sample mean following a 10% increase in GP supply is 0.016 EQ-5D units. If the health status change is sustained over

1 year, the QALY gain is 0.016 QALYs per person. The mean annual cost to the NHS of an additional GP is £170,000 (Netten and Curtis, 2003). This figure includes net remuneration costs (comprising 40% of the total), practice staff costs assuming 0.4 WTE practice nurse and 0.06 WTE other staff per WTE GP (8%), travel costs (2%), other practice expenses (27%), equivalent annual costs of pre- and post-registration training (14%), ongoing training (1%), premises and equipment costs (5%) and overheads (3%). At the sample mean of 0.5 GPs per 1,000 population a 10% increase in GP supply would result in the addition of 0.05 GPs per 1,000 population, or 0.1 GP for every 2,000 people. The incremental cost per person of the 10% increase would therefore be £170,000*0.1/2000 = £8.50. On this basis the incremental cost per QALY gained of a 10% increase in GP supply would be £531 (=£8.50/0.016). The incremental cost per QALY gained following a 10% increase in GP supply ranges from £230 to £531 depending on the EQ-5D based health measure and the GP supply measure.

6 Concluding remarks

In this paper we analysed the impact of GP supply on individual health status using a number of measures. As shown in Table 4 single-equation models that do not control for endogeneity yield insignificant estimates of the impact of GP supply on health status. Using IV methods to control for the endogeneity of GP supply and health we find that GP supply has a statistically significant and positive impact on health status. Thus, failure to allow for endogeneity will yield underestimates of the effect of GP supply on health status.

The analysis in the paper is a first look at the data. Plans for further work include:

- Allowing the effect of GP supply to be non-linear
- Use of the detailed health measures in HSE. For example it is plausible that GPs have different effects on chronic conditions than on acute conditions
- Use of measures of GP quality, such as the training status of practices
- Measures of the quality of care received by individuals derived by examining the drugs prescribed for patients with particular conditions
- Analysing the impact of GP supply on mortality

Further suggestions are very welcome.

Acknowledgements

NPCRDC receives funding from the Department of Health. MS is funded by the Scottish Higher Education Funding Council through the *Platform Project*. The views expressed are not necessarily those of the funders. We are grateful to Madhavi Bajekal and Bob Erens from the National Centre for Social Research, and Andrew Wagner and Mark Hann from NPCRDC for assistance with data.

References

Bunker J. Medicine matters after all: measuring the benefits of medical care, a healthy lifestyle, and a just social environment. Nuffield Trust Series 15. 1995.

Capewell S., Morrison, C.E., and McMurray, J.J. Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994. *Heart*, Apr 1999; 81: 380 - 386.

Cutler D, McLellan M. Productivity Change in Health Care, *American Economic Review*, May 2001, 281-286.

Davey Smith G, Frankel S, Ebrahim S. Rationing for health equity: is it necessary? Health Econ. 9: 575–579 (2000)

Dusheiko, M., Gravelle, H. and Campbell, S. Does primary care make a difference? The relationship between individual health and practice characteristics. September 2003.

Gravelle H, Sutton M. Trends in geographical inequalities in provision of General Practitioners in England and Wales. *Lancet* 1998; 352: 1910 [letter].
Gravelle H, Sutton M. Geographical inequalities in provision of General Practitioners. *Lancet* 1999; 353: 931-2 [author's reply].
Gravelle H. Sutton M. Income-related inequalities in self-assessed health in Britain: 1979-1995. *Journal of Epidemiology and Community Health* 2003; 57:125–129.
Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C., Muirhead, M. A model of supply and demand influences on the use of health care: implications for deriving a 'needs-based' capitation formula. *Health Economics*. 2003, 12, 985-1004.
Guilford, C. Availability of primary care doctors and population health in England: is there an association? *Journal of Public Health Medicine*. 2002, 24, 252-254.l
Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251-71.
Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, Hurwitz B, Iezzoni LI. Explaining differences in English hospital death rates using routinely collected data. *BMJ*, Jun 1999; 318: 1515-1520.
Macinko, J., Starfield, B. and Shi, L. The contributions of primary care systems to health outcomes within Organization for Economic Cooperation and Development (OECD) Countries, 1970-1998. Health Services Research. 2003, 38, 831-865.
McKeown T. *The role of medicine: dream, mirage or nemesis.* London: Nuffield Principal Hospital Trust, 1976.
Morris, S., Sutton, M. and Gravelle, H. (2003). Inequity and inequality in the use of health care in England: an extended empirical investigation. *Centre for Health Economics, Technical Paper*, No 27. University of York.
Netten, A. and Curtis, L. (2003). Unit costs of health and social care. PSSRU, University of Kent.
Sargan, J., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26, 393-415.
Shi, L., and Starfield, B. Primary care, income inequality, and self-rated health in the Untied States: a mixed-level analysis. *International Journal of Health Services Research*. 2000, 30, 541-555.
Shi, L., and Starfield, B. The effect of primary care physician supply and income inequality on mortality among blacks and whites in US metropolitan areas. *American Journal of Public Health*. 2001, 91, 1246-1250.
Shi, L., Starfield, B., Kennedy, B. and Kawachi, I. Income inequality, primary care and health indicators. *The Journal of Family Practice*, 1999, 48, 275-284.
Shi, L., Starfield, B., Politzer, R., and Regan, J. Primary care, self-rated health, and reductions in social disparities in health. *Health Services Research*. 2002, 37, 529-550.
Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557-86.
Starfield, B. New paradigms for quality in primary care. *British Journal of General Practice*. 2001, 51, 303-309.
Starfield, B. Improving equity in health: a research agenda. *International Journal of Health Services*. 2001, 31, 545-566.
Starfield, B. and Shi, L. Policy relevant determinants of health: an international perspective. *Health Policy*. 2002, 60, 201-218.
Sutton M., Gravelle H., Morris S., Leyland A., Windmeijer F., Dibben C., Muirhead M. (2002). *Allocation of Resources to English Areas: Individual and Small Area Determinants of Morbidity and Use of Health Care*. Report for Department of Health Information and Statistics Division, Common Services Agency, Scotland. [http://www.isdscotland.org/isd/info3.jsp?pContentID=2859&p_applic=CCC&p_service=Content.show&]
Sutton M. Vertical and horizontal aspects of socio-economic inequity in General Practitioner contacts in Scotland. *Health Economics* 2002; 11: 537-549.
Tunstall-Pedoe H, Vanuzzo D, Hobbs M, et al. Estimation of contribution of changes in coronary care to improving survival, event rates, and coronary heart disease mortality across the WHO MONICA Project populations. *Lancet* 2000; 355: 688–700
Wooldridge, J.M., 2002. Econometric analysis of cross-section and panel data. MIT Press, Cambridge, Massachusetts.
Wu, D.M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41, 733-750.

Table 1. Summary statistics for health status variables

	Mean	Std. Dev.	Min.	Max.
Whole sample (n=85,555 for h^2, n=86,851 for h^3 and h^5)				
h^2	0.85	0.13	0.21	0.94
h^4	0.82	0.16	0.00	1.00
h^5	0.84	0.16	-0.12	0.99
Self-assessed general health = "very good" (n=32,318 [37.8% of sample with non-missing values])				
h^2	0.94	0.00	0.94	0.94
h^4	0.91	0.08	0.29	1.00
h^5	0.91	0.07	0.23	0.99
Self-assessed general health = "good" (n=34,475 [40.3%])				
h^2	0.88	0.00	0.88	0.88
h^4	0.85	0.11	0.00	1.00
h^5	0.87	0.11	-0.12	0.99
Self-assessed general health = "fair" (n=14,253 [16.6%])				
h^2	0.72	0.00	0.72	0.72
h^4	0.70	0.16	0.10	1.00
h^5	0.72	0.17	-0.07	0.99
Self-assessed general health = "bad" (n=3,505 [4.1%])				
h^2	0.43	0.00	0.43	0.43
h^4	0.53	0.16	0.00	1.00
h^5	0.52	0.19	-0.11	0.97
Self-assessed general health = "very bad" (n=1,004 [1.2%])				
h^2	0.21	0.00	0.21	0.21
h^4	0.45	0.15	0.03	0.96
h^5	0.43	0.20	-0.08	0.96
Self-assessed general health = missing (n=0 for h^2, n=1,296 for h^3 and h^5)				
h^2	-	-	-	-
h^4	0.35	0.09	0.03	0.95
h^5	0.65	0.10	0.32	0.94

Table 2. Summary statistics for GP supply variables (GPs per 1,000 population) (n=66,484)

	g^1	g^2	g^3
Distributions			
Mean	0.50	0.50	0.50
Median	0.50	0.50	0.50
Minimum	0.33	0.33	0.39
5th percentile	0.43	0.43	0.45
95th percentile	0.60	0.59	0.57
Maximum	2.20	2.21	0.68
Std. Dev.	0.07	0.06	0.04
Range	1.87	1.88	0.29
Interquartile range	0.06	0.06	0.05
Correlation coefficients			
g^2	0.96*	1.00	
g^3	0.51*	0.53*	1.00

* p<0.0001

Table 3. Advantages and disadvantages of health status and GP supply variables

Health measures	
h^1	Precludes use of 2SLS estimation. Any reporting bias in self-assessed general health more likely to impact. Less efficient estimation since the cut points of the distribution of the latent variable are estimated
h^2	Coefficient can be interpreted as changes in quality of life suitable for QALY calculations Assumes all individuals in the same self-assessed general health category have same EQ-5D score. Any reporting bias in self-assessed general health more likely to impact
h^3	Coefficient can be interpreted as changes in quality of life suitable for QALY calculations Precludes use of 2SLS estimation. Assumes stable mapping of 1996 EQ-5D scores to self-assessed general health over time. Any reporting bias in self-assessed general health more likely to impact
h^4	Less efficient estimation since the cut points of the distribution of the latent variable are estimated. Residuals from first-stage health equation assumed to be measurement error only
h^5	Coefficient can be interpreted as changes in quality of life suitable for QALY calculations Assumes stable mapping of 1996 EQ-5D scores to self-assessed general health over time. Residuals from first-stage health equation assumed to be measurement error only.
GP supply measures	
g^1	Assumes proportion of people in each ward registered to each practice is constant over time. Mismatch in datasets, which becomes more pronounced over time.
g^2	Assumes proportion of people in each ward registered to practices located in each ward is constant over time. Initial mismatch in datasets.
g^3	Less precise estimates since there is less variation in the supply variable. Ignores within LA variation.

Table 4. Single equation estimates of the impact of GP supply on health

GP supply measure:	g^1		g^2		g^3	
h^1 (Ordered probit)	Coef.	z	Coef.	z	Coef.	z
GP supply, *g*	-0.051	-0.7	-0.041	-0.6	0.134	0.9
N	65,190		65,190		65,190	
Pseudo-R^2	0.08		0.08		0.08	
h^2 (OLS)	Coef.	t	Coef.	t	Coef.	t
GP supply, *g*	-0.002	-0.2	-0.002	-0.2	0.010	0.6
N	65,190		65,190		65,190	
R^2	0.17		0.17		0.17	
h^3 (Interval regression)	Coef.	z	Coef.	z	Coef.	z
GP supply, *g*	-0.005	-0.4	-0.005	-0.4	0.018	0.7
N	65,190		65,190		65,190	
Pseudo-R^2	0.15		0.15		0.15	
h^4 (OLS)	Coef.	t	Coef.	t	Coef.	t
GP supply, *g*	0.007	0.8	0.010	1.2	0.025	1.4
N	66,484		66,484		66,484	
R^2	0.41		0.41		0.41	
h^5 (OLS)	Coef.	t	Coef.	t	Coef.	t
GP supply, *g*	0.007	0.9	0.011	1.2	0.022	1.2
N	66,484		66,484		66,484	
R^2	0.27		0.27		0.27	

The following covariates are included but not reported: age and sex variables; income; social class of head of household; highest educational qualification attained; ethnicity; rurality; number of cars owned by household; marital status; housing tenure; year; month of interview; proxy respondent; region of residence; hospital distance; hospital beds; and, hospital waiting times..

Table 5. IV estimates of the impact of GP supply on health

	g^1		g^2		g^3	
h^1 (IV ordered probit)	Coef.	z*	Coef.	z*	Coef.	z*
GP supply, *g*	1.723	1.8	1.485	1.4	2.480	1.3
N	65,190		65,190		65,190	
F test for instruments=0 [p value]	213.89 [<0.01]		182.72 [<0.01]		280.97 [<0.01]	
Hausman z test for endogeneity*	-1.8		-1.4		-1.2	
h^2 (2SLS)	Coef.	z	Coef.	z	Coef.	z
GP supply, *g*	0.294	3.0	0.281	2.6	0.478	2.5
N	65,190		65,190		65,190	
F test for instruments=0 [p value]	213.89 [<0.01]		182.72 [<0.01]		280.97 [<0.01]	
Sargan θ^2 statistic [p value]	6.86 [0.01]		9.20 [<0.01]		9.60 [<0.01]	
Hausman F test for endogeneity [p value]	9.27 [<0.01]		6.86 [<0.01]		6.18 [0.01]	
h^3 (IV interval regression)	Coef.	z*	Coef.	z*	Coef.	z*
GP supply, *g*	0.463	2.8	0.434	2.3	0.736	2.3
N	65,190		65,190		65,190	
F test for instruments=0 [p value]	213.89 [<0.01]		182.72 [<0.01]		280.97 [<0.01]	
Hausman z test for endogeneity*	-2.8		-2.4		-2.2	
h^4 (2SLS)	Coef.	z	Coef.	z	Coef.	z
GP supply, *g*	0.321	3.2	0.339	3.0	0.591	3.0
N	66,484		66,484		66,484	
F test for instruments=0 [p value]	223.76 [<0.01]		188.80 [<0.01]		276.23 [<0.01]	
Sargan θ^2 statistic [p value]	0.83 [0.36]		1.87 [0.17]		2.28 [0.13]	
Hausman F test for endogeneity [p value]	9.94 [<0.01]		8.76 [<0.01]		8.19 [<0.01]	
h^5 (2SLS)	Coef.	z	Coef.	z	Coef.	z
GP supply, *g*	0.324	3.1	0.351	3.0	0.617	3.0
N	66,484		66,484		66,484	
F test for instruments=0 [p value]	223.76 [<0.01]		188.80 [<0.01]		276.23 [<0.01]	
Sargan θ^2 statistic [p value]	0.14 [0.71]		0.68 [0.41]		0.93 [0.34]	
Hausman F test for endogeneity [p value]	9.35[<0.01]		8.67 [<0.01]		8.36 [<0.01]	

The covariates are the same as in Table 3. *The standard error used to compute the z score is the standard deviation of the coefficient from 50 replications.

Appendix 1. Regression of self-assessed general health on health indicators to predict h^4 and h^5

	Ordered probit (h^4)		Interval regression (h^5)	
	Coef.	z	Coef.	z
Limiting longstanding illness	-0.638	-46.2	-0.130	-63.6
Days cut down due to acute ill health [1]				
1 to 3 days	-0.190	-11.2	-0.027	-10.9
4 to 6 days	-0.324	-13.6	-0.053	-14.9
7 to 13 days	-0.412	-17.1	-0.077	-21.1
14 days	-0.531	-28.8	-0.131	-46.3
Specific longstanding illnesses				
Neoplasms	-0.609	-16.6	-0.085	-15.1
Endocrine and metabolic	-0.586	-26.4	-0.069	-20.4
Mental disorder	-0.428	-15.5	-0.027	-6.5
Nervous system	-0.432	-18.2	-0.033	-9.2
Eye	-0.173	-6.0	0.014	3.4
Ear	-0.224	-8.3	0.011	2.7
Heart and circulatory	-0.633	-35.1	-0.080	-29.6
Respiratory	-0.630	-42.4	-0.066	-30.3
Digestive	-0.545	-24.3	-0.058	-17.0
Genitourinary	-0.484	-16.0	-0.047	-10.3
Skin	-0.282	-10.8	-0.003	-0.8
Musculoskeletal	-0.338	-21.2	-0.020	-8.6
Infectious disease	-0.651	-7.6	-0.074	-5.6
Blood disorders	-0.585	-10.9	-0.073	-8.8
Other longstanding illness	-0.547	-6.1	-0.055	-4.0
Number of longstanding illness [2]				
2	0.108	5.3	-0.036	-11.8
3	0.283	8.3	-0.083	-16.1
4 or more	0.604	11.2	-0.114	-14.0
GHQ-12 score [3]				
1	-0.204	-13.3	-0.024	-11.1
2	-0.270	-14.0	-0.038	-13.5
3	-0.315	-13.4	-0.045	-13.1
4	-0.405	-14.7	-0.064	-15.6
5	-0.473	-15.3	-0.083	-18.0
6	-0.503	-14.0	-0.087	-16.3
7	-0.568	-14.7	-0.100	-17.1
8	-0.593	-13.7	-0.113	-17.2
9	-0.581	-12.6	-0.108	-15.2
10	-0.654	-13.8	-0.128	-17.5
11	-0.646	-12.6	-0.129	-16.3
12	-0.727	-14.0	-0.156	-19.5
Health-related economic inactivity [4]				
Retired	-0.100	-6.4	-0.018	-8.0
Permanently sick	-0.896	-37.1	-0.243	-64.6
Temporary sickness	-0.640	-9.0	-0.135	-12.0
N	85,555		85,555	
Pseudo-R^2	0.18		0.46	

Both models also include age and sex variables and HSE year variables. [1] The baseline category is zero days. [2] The baseline category is 0 or 1. [3] The baseline category is 0. [4] The baseline category is economically active or inactive for non-health reasons.

Appendix 2. IV estimates of the impact of GP supply measure g^1 plus covariates on health measures h^3 and h^5

	h^3 (IV interval regression)				h^5 (2SLS)			
	GP supply equation		Health equation		GP supply equation		Health equation	
	Coef.	t	Coef.	z*	Coef.	t	Coef.	z
GP supply, g^1			0.463	2.8			0.324	3.1
Instruments								
Proportion population aged 75+	0.518	5.8			0.502	5.7		
Age weighted capitation payment	-0.010	-3.5			-0.010	-3.3		
Income/100,000	0.003	1.5	0.059	12.5	0.003	1.5	0.046	12.4
Social class of head of household [1]								
II Managerial/technical	0.002	1.8	-0.002	-0.7	0.002	1.8	-0.003	-1.1
IIIn Skilled non-manual	-0.001	-0.5	0.002	0.5	-0.001	-0.5	0.003	1.0
IIIm Skilled manual	0.000	-0.3	-0.018	-5.1	0.000	-0.4	-0.003	-1.0
IV Semi-skilled manual	0.001	0.6	-0.020	-5.3	0.001	0.5	-0.003	-1.1
V Unskilled manual	0.000	-0.1	-0.033	-5.1	0.000	-0.1	-0.002	-0.5
Other	-0.001	-0.3	-0.019	-3.6	-0.001	-0.4	-0.001	-0.1
Highest educational qualification [2]								
Higher education less than a degree	0.002	1.7	-0.015	-3.8	0.002	1.7	-0.005	-1.9
A level or equivalent	-0.002	-1.5	-0.015	-4.9	-0.002	-1.5	-0.009	-3.3
GCSE or equivalent	-0.001	-0.6	-0.023	-7.4	-0.001	-0.7	-0.007	-3.1
CSE or equivalent	0.000	-0.1	-0.047	-8.6	0.000	-0.1	-0.016	-4.7
Other qualification	-0.002	-0.9	-0.035	-5.4	-0.001	-0.9	-0.018	-4.9
No qualification	-0.003	-2.5	-0.083	-22.3	-0.003	-2.4	-0.034	-14.0
Ethnic group [3]								
Black Caribbean	0.006	3.8	-0.025	-5.4	0.006	3.9	-0.004	-1.1
Black African	0.004	1.3	0.005	0.5	0.004	1.4	0.005	0.7
Black Other	-0.004	-1.2	-0.032	-3.1	-0.004	-1.2	0.000	-0.1
Indian	-0.002	-1.7	-0.052	-11.4	-0.002	-1.7	-0.007	-2.4
Pakistani	0.001	0.9	-0.055	-12.1	0.001	0.9	-0.004	-1.4
Bangladeshi	0.024	14.7	-0.073	-9.6	0.024	14.9	0.004	0.9
Chinese	0.026	12.3	-0.047	-5.4	0.026	12.4	0.024	4.3
Other non-white ethnic group	0.011	6.2	-0.027	-4.3	0.011	6.2	-0.007	-1.8
Rurality [4]								
Suburban	-0.020	-29.1	-0.004	-1.0	-0.020	-29.2	-0.001	-0.2
Urban	-0.018	-19.1	-0.006	-1.4	-0.018	-19.5	0.003	1.1
Number of cars owned by household [5]								
1	-0.001	-1.6	0.028	11.8	-0.001	-1.6	0.018	11.2
2	0.001	1.5	0.032	11.4	0.001	1.6	0.023	11.5
3 or more	0.000	-0.2	0.033	10.9	0.000	-0.1	0.023	7.9
Marital status [6]								
Married	-0.001	-1.4	0.016	5.9	-0.001	-1.4	0.025	12.8
Separated	0.000	-0.2	-0.004	-0.5	0.000	-0.2	-0.003	-0.8
Divorced	-0.001	-0.8	-0.011	-1.9	-0.001	-0.8	-0.011	-3.6
Widowed	-0.003	-1.8	0.015	2.1	-0.003	-1.8	0.013	3.9
Housing tenure [7]								
Mortgage	0.002	2.6	-0.019	-7.5	0.002	2.5	-0.005	-3.3
Part mortgage part rent	0.001	0.4	-0.068	-5.3	0.001	0.4	-0.015	-2.1
Rent	0.007	9.0	-0.053	-17.5	0.007	9.1	-0.028	-14.1
Living rent free	0.008	2.9	-0.014	-1.5	0.007	2.9	-0.009	-1.6
Hospital supply variables								

Distance to 5 nearest acute providers/100	0.107	35.9	-0.063	-2.9	0.107	36.2	-0.019	-1.4
Beds at acute providers used/1000	0.011	12.5	-0.005	-1.6	0.010	11.9	-0.007	-3.1
Patients waiting less than 6 months	0.014	3.3	-0.012	-1.0	0.018	4.4	-0.016	-1.7
N	65,190				66,484			
F test for instruments=0 [p value]	213.89 [<0.01]				223.76 [<0.01]			
Sargan θ^2 statistic [p value]					0.14 [0.71]			
Hausman-type z test for endogeneity*	-2.8							
Hausman F test for endogeneity [p value]					9.35[<0.01]			

The models also include age and sex variables, HSE year dummies, item non-response dummies, GOR dummies, month of interview dummies, and a proxy respondent dummy. *The standard error used to compute the z score is the standard deviation of the coefficient from 50 replications. [1] The baseline category is I. [2] The baseline category is Degree. [3] The baseline category is White. [4] The baseline category is Rural. [5] The baseline category is 0. [6] The baseline category is Single. [7] The baseline category is Own outright.