HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 11/22

Waiting times and socioeconomic status: does sample selection matter?

Anurag Sharma
Luigi Siciliani
Anthony Harris

July 2011

york.ac.uk/res/herc/hedgwp

Waiting times and socioeconomic status: does sample selection matter?

Anurag Sharma[1] Luigi Siciliani[2] Anthony Harris[3]

27 July 2011

Abstract

An increasing amount of empirical evidence suggests that patients with higher socioeconomic status wait less within publicly-funded hospitals to receive non-emergency (elective) surgery. Using data from Australia, we investigate the extent to which such gradient can be explained by sample selection, with richer patients being more likely to opt for treatment in the private sector when faced with waiting times in the public sector. We show that, once the potential biases introduced by sample selection are taken into account, the gradient between waiting times and socioeconomic status reduces significantly in size but does not disappear.

Keywords: Hospital Waiting Times; Socio-economic Gradient; Quantile regression with sample selection; Heckman model

JEL Classification: I10;C14;C21

[1] *Corresponding author*. Centre for Health Economics, Monash University, Clayton Victoria 3800, Australia. E-mail anurag.sharma@monash.edu.

[2] Department of Economics and Related Studies, and Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK; and C.E.P.R., 90-98 Goswell Street, London EC1V 7DB, UK. E-mail: luigi.siciliani@york.ac.uk.

[3] Centre for Health Economics, Monash University, Clayton, Victoria 3800, Australia. E-mail: anthony.harris@monash.edu.

1. Introduction

Many OECD countries provide universal access to required health care, irrespective of the ability to pay. For example, in the UK and Australia, all permanent residents have the right to public hospital treatment at no charge. An important aim of such systems is to provide care according to clinical need and not on the basis of willingness to pay.

Publicly-funded systems are however often characterised by excess demand which generates a waiting list. Waiting times generate dissatisfaction for patients since they postpone benefits from treatment, may induce a deterioration of the health status of the patient, prolong suffering and generate uncertainty. Given the absence of a monetary price, economists have argued that waiting times act as a non-monetary price which brings the demand for and supply of health care in equilibrium (Lindsay and Feigenbaum, 1984; Martin and Smith, 1999; 2003; Martin et al., 2007; Cullis, Jones and Propper, 2000; Fabbri and Monfardini, 2009).

One apparent advantage of rationing by waiting times is that access to services does not depend on the ability to pay, unlike any form of price rationing where access is dependent on income. However, a recent empirical literature (reviewed in more detail below) suggests that, even within a publicly funded health system, non-price rationing does not guarantee equality of access by socio-economic status. Individuals with higher socioeconomic status (as measured by income or educational attainment) can wait less for publicly-funded hospital care than those with lower socioeconomic status (Siciliani and Verzulli, 2009; Cooper et al., 2009; Laudicella et al., 2010; Johar et al., 2010; Monstad, 2010; Carlsen and Kaarboe, 2010). This negative gradient between waiting time and socioeconomic status may be interpreted as evidence of inequity within publicly-funded systems which favours rich and more educated patients over poorer and less educated ones. Therefore, rationing by waiting times may be less equitable than it appears in first instance.

Publicly-funded healthcare systems are often characterised by the presence of a private sector alternative which patients can access if they pay out of pocket or if they are privately insured. A key feature of this private sector is that waiting times are negligible: the promise of low wait is indeed the main way to attract patients from the public to the private sector (Iversen, 1997; Hoel and Saether, 2003; Olivella, 2003; Marchand and Schroyen, 2005). Given such institutional feature, one possible explanation for the observed gradient between waiting time and socioeconomic status is the possibility of sample selection: rich patients who expect high waiting times are more likely to afford and opt for the private sector generating a negative gradient between income and waiting time in the public system. In other words, public hospitals treat poor patients with expected high and low waiting times but only rich patients with low waiting times.

The objective of this paper is to explore the role of sample selection generated by the presence of a private sector in explaining the gradient between waiting time and socioeconomic status. We use data from Australia and exploit the unique feature of administrative databases that collect extensive information for both public and private patients. We estimate Heckman sample selection model which explicitly takes into account the choice of public versus private care. This is in contrast to previous studies where administrative databases only collect data for public patients (this is the case in England and Norway for example): without data from the private sector Heckman sample selection models cannot be tested and the scope of sample selection in explaining the gradient is difficult to evaluate. The investigation of the role of sample selection into explaining the gradient is our key contribution to this recent literature. As far as the authors are aware, this is the first paper which tackles sample selection in the estimation of the gradient between waiting times and socioeconomic status.

We use data on over 200,000 elective hospital admissions from the state of Victoria in Australia in the financial year 2005-06, where the private sector accounts for half of all hospital admissions: the potential bias due to sample selection may therefore be a significant one. Comparable data are available at patient level for both the public and the private sector.

We investigate whether patients living in better-off areas wait less in public hospitals than patients in poorer areas. Socioeconomic status is measured through an index which captures economic resources at small area level: Socio-Economic Indexes for Areas (SEIFA) for Economic Resources. Waiting times refer to the time elapsed between the time the specialist adds the patient to the waiting list and the time the patient is admitted for hospital treatment (also known as 'inpatient' waiting time). Our sample includes all patients in need of elective surgery.

We first run a baseline OLS regression of waiting times in public hospitals on socioeconomic status and compare it with the estimates from a Heckman sample selection model where the first stage is modelled as a choice between public and private hospital treatment and the second stage provides an estimate of the waiting time-income gradient once corrected for possible sample selection.

We also expand the analysis using quantile regression to explore the gradient and sample selection at five quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th}). This is because we expect sample selection to matter more at high levels of waiting times when the incentive to opt for the private sector is strongest: patients expecting high waiting times are more likely to opt for private hospital care than those expecting lower ones. The quantile estimates control for sample selection using the semi-parametric correction suggested by Buschinsky (1998).

The analysis overall suggests that sample selection plays a role in explaining the gradient. The gradient is smaller after controlling for sample selection but does not vanish. Using a sample of all patients in need of non-emergency surgery, OLS regression shows a negative waiting time-gradient: compared to patients coming from the poorest areas, those coming from the richest ones wait about 12% less. With an average waiting time of 89 days, this amounts to over a week. Since we control for hospital fixed effects, the gradient can be interpreted as evidence of inequalities within the hospital. However, once the estimates are corrected for sample selection using the Heckman approach, the magnitude of the gradient reduces by about half.

The first-stage selection equation confirms that patients residing in the richest areas have a lower probability of choosing a public hospital. For example, relative to patients residing in the poorest areas, patients in need of surgery residing in areas in the third decile of deprivation are 26% less likely to opt for a public hospital, and those in the richest decile are 80% less likely to opt for a public hospital.

Quantile regression (without correcting for sample selection) suggests that compared to patients living in areas with lowest income, patients living in areas with income in the top three deciles wait 8-12% less across different quantiles of the waiting time distribution. Similarly, patients living in areas with income in top 6-7 deciles wait 3.3-7.7% less compared to patients living in areas with income in the lowest decile across all quantiles of the waiting time distribution. In line with previous results, once sample selection is taken into account, the gradient is substantially reduced.

We also replicate the analysis for four sub-groups of patients in need of Eye Surgery, Hip & Knee Surgery, Prostatectomy and Hysterectomy. The results show that whenever a gradient is detected by OLS, this is reduced once sample selection is taken into account.

1.1 Existing literature

In this section we review the existing literature. Using survey data from SHARE, Siciliani and Verzulli (2009) find that for specialist consultation, individuals with high education experience a reduction in waiting times of 68% in Spain, 67% in Italy and 34% in France (compared to those with low education). Individuals with intermediate education report a waiting-time reduction of 74% in Greece. For non-emergency surgery, they find evidence of a negative and significant association between education and waiting times in Denmark, the Netherlands and Sweden. High education reduces waiting by 66, 32 and 48%, respectively. They also find the presence of income effects, although generally modest. An increase in income of 10,000 Euro reduces waiting times for specialist consultation by 8% in Germany and waiting times for non-emergency surgery by 26% in Greece. Surprisingly, an increase in income of 10,000 Euro increases waiting by 11% in Sweden.

Using administrative data, Cooper et al. (2009) investigate whether in England waiting time inequalities for some elective procedures (i.e. hip and knee replacement, and cataract surgery) varied during the Labour government between 1997 and 2007. While in 1997 waiting times and deprivation tended to be positively related, by 2007 the relation between deprivation and waiting time was less pronounced. They therefore conclude that reforms introduced by the Labour government in terms of patient choice, provider competition, and higher capacity did not conflict with equity.

Laudicella et al. (2010) use administrative data from England to investigate whether patients with higher socioeconomic status (as measured by small area level income and education deprivation) wait less than other patients. The analysis focuses on hip replacement only in 2001. They provide evidence of inequity in waiting times favouring more educated individuals and, to a lesser extent, richer individuals. Patients in the third-to-fifth education-deprived quintile wait about 14% longer (32 days). Patients in the fourth and fifth most income-deprived quintiles wait about 7% longer (18 days) than patients in the least deprived quintile. Since they control for hospitals' fixed effects, the results can be interpreted as evidence of inequalities within the hospital (as opposed to across hospitals).

Johar et al. (2010) use administrative data from New South Wales in Australia to decompose variations in waiting times that are due to clinical need and non-clinical factors such as socioeconomic status and location. They find waiting times to be influenced by non-clinical factors. Discrimination is in favour of the most wealth-advantaged groups with delays of 2 to 3 months at the very top end of the waiting time distribution, and in favour of patients living in more remote areas.

Using administrative data for patients in need of hip replacement in Norway, Monstad et al. (2010) find that a statistically highly significant negative association between income and waiting time for men, and between education and waiting time for women. Carlsen and Kaarboe (2010) use administrative data from all somatic elective inpatient and outpatient hospital stays in Norway. They find little indication of discrimination with regard to income but some indication of discrimination of men with low education in terms of lower probability of zero waiting. They also find a pro-educational bias for women: women with only primary education wait about 9% (13%) longer than women with upper secondary (tertiary) education.

Using administrative data for Sweden, Tinghög et al. (2010) find that individuals with low disposable income experience longer waiting times for orthopedic (27%) and general surgery (34%). No significant differences were found on the basis of gender and ethnicity.

None of the above papers deals with the potential sample selection generated by the presence of a private sector. This study also contributes to the broader literature of measuring equity in healthcare utilisation (Van Doorslaer and Wagstaff, 2000), which tests whether individuals with higher socioeconomic status have higher utilisation of healthcare, controlling for need, within publicly-funded health systems.

The outline of the study is as follows. Section 2 presents the econometric specification. Section 3 describes the data and institutional setting. Section 4 provides the results and Section 5 concludes.

2. Econometric specification

2.1 OLS model

We are interested in testing whether public patients with higher socioeconomic status wait less than patients with lower socioeconomic status when admitted to hospital. Define *w* as the waiting time between the time the public patient is registered to the waiting list, after specialist assessment, and the time the patient is admitted to a public hospital for treatment. The model specification is the following:

$$\ln(w_{ijk}) = h_j + g_k + \boldsymbol{\beta}_1' y_{ijk} + \boldsymbol{\beta}_2' p_{ijk} + u_{1ijk} \qquad (1)$$

where w_{ijk} is the waiting time of patient *i* in public hospital *j* which receives a treatment with DRG *k*; y_{ijk} is a vector of dummy variables which captures socioeconomic status, as measured by the income in the area where the patient resides (across ten quantiles of the income distribution). Inequalities in waiting time across patients with different socioeconomic status arise if $\boldsymbol{\beta}_1 \neq \mathbf{0}$. Suppose that the reference group includes patients with lowest socioeconomic status. Then, a negative gradient implies that patients residing in wealthier areas wait less.

p_{ijk} is a vector of patients' characteristics. Patients' characteristics include age, gender, total number of procedures performed during a hospital's episode and primary diagnosis. These factors control for severity of patient's health condition. Patients registered on the waiting list are prioritised on the basis of their severity and more severe patients wait less relative to non-severe ones (Gravelle and Siciliani, 2008a). Moreover, more severe patients (with worse health) are more likely to have lower socioeconomic status (Wagstaff and van Dooerslaer, 2000). Thus patients' severity might be correlated negatively with both waiting time and socioeconomic status. Failure to control for patient severity might generate biased results: for example, without controlling for severity there may be a positive correlation between waiting and income while such correlation disappears once controls for severity are added.

g_k is a DRG fixed effect which captures differences in waiting times across different treatments. For example waiting times for coronary bypass are systematically shorter than for cataract surgery which reflects prioritisation across different treatments (Gravelle and Siciliani, 2008b; see Siciliani and Hurst, 2004 for international empirical evidence).

Hospital fixed effects h_j are included in the model to control for differences in waiting times across hospitals which might arise from differences in supply (beds, doctors, efficiency). Hospitals with high supply (and lower waiting times) might be located in urban areas where high-income patients are concentrated leading to correlation between hospital characteristics and socio-economic characteristics of patient's area of residence. In this scenario, omitting hospital fixed effects might overestimate inequalities. Including hospital fixed effects also allows interpreting socioeconomic inequalities in waiting times "within" a hospital, rather than across hospitals.

The dependent variable w_{ijk} is transformed by the logarithmic function to reduce the skewness of the waiting time distribution. u_{ijk} is the idiosyncratic error. We estimate Equation (1) with OLS.

Equation (1) refers to patients receiving treatment in public hospitals only. It therefore does not include patients who opted for private hospital care. Estimating Equation (1) by OLS may introduce a sample selection bias if richer patients who expect high waiting times in public hospitals opt for private hospital care therefore leaving poor patients in public hospitals waiting longer.

OLS may therefore overestimate the waiting-time gradient in public hospitals if socioeconomic status is positively correlated with the probability of going private. In other words, OLS may find that richer patients wait less in public hospitals, but this is explained more by a reduction in their waiting times as they substitute into private system when their expected waiting times in the public system are high, than by a direct effect of income on waiting times in public hospitals.

One way to address this issue is to make use of the sample selection model proposed by Heckman (Heckman 1979), which is described below.

2.2 Heckman Selection Model

Define s_{ik} as a dummy variable which takes the value 1 if patient *i* in DRG k seeks care in a public hospital and 0 if the patient seeks care in a private hospital. We define the selection equation as follows:

$$s_{ik} = Z_{ik}\boldsymbol{\beta}_{\boldsymbol{sel}} + u_{2ik} \qquad (2)$$

where Z_{ik} includes dummy variables which captures socioeconomic status (nine dummies for ten quantiles, patient characteristics (age, gender, total number of procedures performed during a hospital's episode and primary diagnosis), DRG fixed effects and patient's private health insurance status.

Since w_{ijk} is observed only if the patient seeks treatment in a public hospital, Equation (1) can be rewritten as:

$$\ln(w_{ijk}) = h_j + g_k + \boldsymbol{\beta}_1' y_{ijk} + \boldsymbol{\beta}_2' p_{ijk} + u_{1ijk} | s_{ik} = 1 \qquad (3)$$

where $u_1 \sim N(0, \sigma)$ and $u_2 \sim N(0,1)$. The covariance between the two error terms is given by $\rho = Cov(u_1, u_2)$. As mentioned above, estimating Equation (1) using OLS in this scenario may lead to biased estimates. Heckman selection method involves a two-step estimation procedure which provides consistent, asymptotically efficient estimates for all parameters. The Probit equation (Equation (2)) is estimated by maximum likelihood to obtain $\widehat{\boldsymbol{\beta}}_{sel}$. For each patient in the selected sample, the Inverse Mill's Ratio (IMR) is computed as $\hat{\lambda}$ = $\varphi(\boldsymbol{Z}'\,\widehat{\boldsymbol{\beta}}_{sel})/\Phi(\boldsymbol{Z}'\,\widehat{\boldsymbol{\beta}}_{sel})$. Equation (3) is subsequently estimated by ordinary least square regression including $\hat{\lambda}$ as an additional covariate. A non-zero significant coefficient of the Inverse Mill's Ratio ($\beta_\lambda = \sigma\rho$) provides evidence of sample selection.

For the identification of the selection model we need to apply at least one exclusion restriction, ie one variable that is included in Z_{ij} but is excluded as a covariate in the waiting time equation (Equation (1)).[4] We therefore require at least one variable that affects the probability of seeking public/private hospital care but not the waiting times in public hospitals. We use patient's private health insurance status as an identifying variable. The main reason for this is that a decision to seek care in a private hospital system is mainly based on patient's private health insurance status and thus will be crucial in the first stage selection equation, while in this particular health system, waiting times for public patients are unlikely to be affected by insurance status.

[4] It should be noted that under the assumption of normality in the probit model, IMR is a nonlinear function of Z and the second stage equation can be identified because of this non-linearity alone even if both equations (2 and 3) have the same set of covariates. However, in practice relying on functional form alone for identification is not recommended (Jones, 2007) and we use the exclusion restriction for identification.

2.3 Quantile Regression with Sample Selection

The Heckman model discussed above corrects for sample selection at the conditional mean. The relationship between waiting times and socioeconomic status, and sample selection bias may however vary at different levels of waiting times. Since patients expecting very high waiting times are more likely to opt for private hospital care, selection bias might be more pronounced at very high levels of waiting times. To account for this we use quantile regression (QR) analysis to investigate the gradient and selection bias across the whole distribution of waiting times. Two main advantages of using a quantile-regression approach are that: i) it uses the entire sample information without losing degrees of freedom which instead occurs if the sample is split into sub-samples; and ii) it avoids the sample truncation problem which arises by splitting the data into sub-samples based on observed values of the waiting times which in itself could cause selection bias.

We can rewrite equation (3) for the τ^{th} conditional quantile (Q_τ) in the following way (Buschinsky, 1998):

$$Q_\tau\big(\ln\big(w_{ijk}\big|s_{ijk}=1\big)\big) = h_j(\tau) + g_k(\tau) + \beta_1^{'}(\tau)y_{ijk} + \beta_2^{'}(\tau)p_{ijk} + Q_\tau(u_{1ijk}(\tau)|s_{ijk}=1) \tag{4}$$

where β_τ are the slope coefficients at the τ^{th} quantile. In the absence of sample selection $Q_\tau\big(u_{1ijk}(\tau)\big|s_{ijk}=1\big)=0$ and the above equation could be estimated consistently by quantile regression. However, the dependence between u_1 and u_2 will lead to nonzero values of $Q_\tau\big(u_{1ijk}(\tau)\big|s_{ijk}=1\big)$. The conditional quantile of observed waiting times depends, among other things, on the quantile-specific selection bias term of unknown form $\mu_\tau(\boldsymbol{Z}'\boldsymbol{\beta}_{sel})$ which cannot be corrected using the traditional IMR derived from the first stage of Heckman model (a fully parametric estimation). Parametric estimators are consistent and asymptotically efficient only if the distributional assumptions are correctly specified. Since in our case the selection bias term is of unknown form, estimating it by parametric method will lead to inconsistent estimates.[5] Therefore, we use a non-parametric method where the unknown function $\mu_\tau(\boldsymbol{Z}'\boldsymbol{\beta}_{sel})$ is approximated by a series expansion, following a two-step procedure. In the first step the selection parameter $\boldsymbol{\beta}_{sel}$ is estimated. In the second step a quantile regression is run on all the second stage covariates and on an estimate of $\mu_\tau(\boldsymbol{Z}'\boldsymbol{\beta}_{sel})$,

[5] For a more technical discussion see Buschinsky (1998).

$\hat{\mu}_\tau(\boldsymbol{Z}'\boldsymbol{\beta}_{sel}) = \mu_\tau(\boldsymbol{Z}'\hat{\boldsymbol{\beta}}_{sel})$. The parameters $\boldsymbol{\beta}_{sel}$ in the selection equation (Equation (2)) are estimated using quasi-maximum likelihood (ML) method following a semi non-parametric (SNP) estimator proposed by Gallant and Nychka (1987). The main reason for using SNP estimator is that it approximates the multivariate probability density (required for the likelihood function) using polynomial expansion (the degree of polynomial determined by a series of Likelihood Ratio (LR) tests) and provides consistent estimates of model parameters. A more detailed discussion on the estimation procedure is provided in Appendix A.

3. Institutional setting and Data

Our sample includes patients from the state of Victoria in Australia. It is the second most populous state and accounts for 25% of the country's population. Patients in the state of Victoria can access elective (non-emergency) treatment either in a public or a private hospital. Patients who seek treatment in a public hospital receive treatment for free under Medicare (Australia's universal public health insurance scheme) but they have to wait.[6] Patients who seek treatment in a private hospital incur the full cost of treatment, which is paid by the patient either directly or through her private health insurer. They do not have to wait for hospital treatment and can choose their own specialist. The private option is clearly attractive for patients who already have a private health insurance which reimburses the cost of hospital treatment, though patients may have a co-payment.

The patient can be referred to a public hospital either from a specialist or a general medical practitioner (a family doctor). Patients who require elective surgery are added to the waiting list. The referring doctor assigns an urgency category based on patient's clinical condition. Patients are informed about the proposed procedure, their provisional diagnosis and their urgency category. Medical practitioners advise patients on the costs associated with treatment in the private sector. They also provide information on access to surgery in public hospital which is based on clinical urgency regardless of insurance status. Elective surgery waiting lists are managed to ensure that priority is given to patients with an urgent clinical need.

[6] Australian Medicare scheme is different than the US Medicare as it covers the whole population and not just the elderly like in the US.

This study uses the Victorian Admitted Episodes Dataset (VAED) for financial year 2005-06. This data is collected for state of Victoria**,** the most densely populated state in Australia where 75% of population lives in the capital city of Melbourne. VAED is an administrative dataset which records each patient admitted to public and private hospitals. It is used for health services planning, policy formulation and funding purposes. The dataset includes over one million of patient episodes each year and records information on length of stay, diagnosis, patient residence, admission type, insurance status, casemix, type of stay and mode of discharge from the hospital.

VAED data for public hospitals includes data on waiting times for patients admitted to elective treatment in public hospitals. Waiting time (which will be our key dependent variable) for each patient is calculated as a difference between the date of registration on the waiting list (from the referring doctor) and the date the patient was admitted for treatment in the public hospital. Private hospitals do not record waiting times because they are on average negligible.

We measure socioeconomic status through an index which captures economic resources at small area level, known as the SEIFA (Socio-Economic Indexes for Areas) for Economic Resources. This index is created from information collected in the Census of Population and Housing in year 2006 (census is conducted every five years).

The following variables are used to generate the SEIFA index for Economic Resources: i) the proportion of people with stated annual household equivalised income between $13,000 and $20,799 (approximately 2nd and 3rd deciles of the income distribution in Australia); ii) the proportion of single-parent families with dependent offspring only; iii) the proportion of occupied private housing with no car; iv) the proportion of households renting from Government or Community organisation; v) the proportion of households paying rent less than $120 per week (excluding $0 per week); vi) the proportion of people aged 15 years and over who are unemployed; vii) the proportion of households who are single households; viii) proportion of occupied private houses requiring one or more extra bedrooms (based on Canadian National Occupancy Standard); ix) the proportion of households owning a house they occupy (without a mortgage); x) the proportion of houses with at least one person who is an owner of an unincorporated enterprise; xi) the proportion of households paying mortgage greater than $2,120 per month; xii) proportion of households owning house (with a mortgage); xiii) proportion of households paying rent greater than $290 per week; xiv) proportion of people with stated annual household equivalised income greater than $52,000 (approx 9th and 10th deciles of the income distribution of Australia); and xv) proportion of occupied private house with four or more bedrooms.

The SEIFA scores are created by combining these variables using principal components analysis (PCA). The aim of PCA is to summarise a large number of correlated variables into a smaller set of transformed variables, called 'principle components'. Each component is a weighted linear combination of the original variables. The first principal component scores for one area are derived by first taking the product of each standardised variable with its respective weight, and then taking the sum across all variables. For convenience of presentation SEIFA scores are standardised such that the mean score in each domain is 1000 with a standard deviation of 100. A low SEIFA score is associated with relative disadvantage.

We use SEIFA deciles[7] reported at the State Suburb Code (SSC) level. SSC is a geographical area created by the Census which aggregates Census Collection Districts (the smallest geographic unit for collection of Census data) to approximate suburbs. In our sample, patients came from 3695 SSCs for treatment in both public and private hospitals in 2005-06. Each SSC has a median population of 781 and an average population of 3340. 75% of SSCs have a population below 3721. We use SEIFA deciles for each SSC which are constructed as follows. All SSCs are ordered from lowest to highest score, then the lowest 10% of SSCs are given a decile number of 1, the second lowest 10% of SSC are given a decile number of 2 and so forth, up to the highest 10% of SSC which are given a decile number of 10. Thus areas are divided into ten groups, depending on their score. Decile 1 is the most disadvantaged relative to the other deciles.

In this analysis we focus on publicly-funded patients receiving treatment in a public hospital and on privately-funded patients receiving treatment in a private hospital. In Victoria there is however a third category of patients, ie patients who are privately-funded but are treated in a public hospital: such private chargeable patients can choose doctor and single-room accommodation if available. This group of patients is however negligible and the diagnosis of these patients are relatively simple from a clinical point of view (for example circumcision, tonsillectomy, inguinal and femoral hernia, myringotomy with tube insertion, ie ear infection etc). We therefore exclude them from the sample.

[7] Since this SEIFA index captures economic resources we use the term income decile interchangeably with SEIFA decile throughout the paper.

We control for patient complexity using the following variables which are reported in VAED and are measured at patient level: age, gender, total number of procedures performed, primary diagnosis, DRG, and whether the patient has private health insurance (PHI). The data for PHI status is part of VAED. It is possible that patients seeking public hospital treatment might not disclose that they have private health insurance. Although this effect will be minimum in our setting as we use PHI (used as an identifying variable) only in the selection equation and not in the waiting time equation, we compare PHI coverage in our combined dataset of public and private hospital patients to that one derived from the National Health Survey (NHS) 2004-05 for the state of Victoria. In the administrative VAED sample the proportion of hospitalised patients with PHI is 44.45%. In the NHS survey sample the PHI figure for hospitalised patients is 43%. These figures are reasonably close and the PHI variable from the administrative data is therefore accurate.

Table 1 provides descriptive statistics for the sample of all patients who received an elective surgery in financial year 2005-6. This includes patients from 275 different DRGs[8] and 4353 primary diagnoses. The number of patients treated in public and private hospitals is respectively 88273 and 131169. Therefore, 60% of the patients go private. There are 25 private hospitals and 30 public hospitals.

The average waiting time in public hospitals is about 89 days. As expected, there are more patients going public from low-income areas and more patients going private from high-income areas. For example, 13% of patients living in the poorest areas choose a public hospital (as opposed to 8% across the whole sample). In contrast, only 2.6% of patients living in richest areas choose a public hospital (as opposed to 5.8% across the whole sample).

About 48% of patients in the whole sample hold private health insurance. The proportion of male patients is 44% in both public and private hospitals. In both types of hospital, most patients receive on average two or three procedures, with only 8-10% receiving more than 5 procedures. Patients' age distribution is similar in public and private hospitals.

Table 1 provides descriptive statistics for all patients who received surgery but also for specific common elective procedures characterised by long waits (Siciliani and Hurst, 2004): Same-Day Eye surgery, Hip and Knee Surgery, Prostatectomy and Hysterectomy. Waiting

[8] The DRG classification system used in the dataset is based on the fourth edition of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10 AM).

times are respectively equal on average to 108, 196, 82 and 95 days. As with the whole sample there are more patients going public from low-income areas and more patients going private from high-income areas.

[Table 1 here]

4. Results

Table 2 provides the results when patients from all DRG groups are included in the sample. The first column provides OLS results on the variation of waiting time in public hospitals across income deciles, after controlling for age, gender, primary diagnosis, DRG, number of procedures and hospital fixed effects.

It suggests that individuals who live in richer areas wait less. More precisely, compared to patients living in areas with lowest income, patients living in areas with highest income wait 13% less. Given an average waiting of 89 days, this implies an average reduction of 11 days. Patients in almost every decile of income have a progressively lower waiting time than the one below. For example, patients living in areas with income falling between the 2nd and the 7th decile wait between 5-8% less than those in the lowest income decile. [9] Patients living in areas with income falling in the 8th or 9th decile wait 10-11% less than those in the lowest (first) income decile.[10]

[Table 2 here]

The second and third columns in Table 2 show the results from the Heckman selection model: the second column provides the waiting time equation and the third column the selection equation. It suggests that the gradient is still present but it is reduced when sample selection is taken into account. Compared to patients in the lowest income decile, patients whose income falls between the 2nd and 7th decile wait 3-4% less, and patients whose income falls between the 8th and 10th decile wait 5-6% less. The precision of the estimate also reduces

[9] Calculated as *exp(β) -1* as waiting time is in logs.

[10] In the model we do not control for education deciles derived from the SEIFA index for education and occupation because of its high collinearity to SEIFA index for Economic Resources. When including both income and education deciles in the OLS and the Heckman selection model, we find no effect of education on waiting times, while the income one remains both statistically significant and of similar magnitude. To simplify the presentation of the results and to improve the efficiency of the estimates we therefore do not include the education variables. Including them does not change in any way the results.

especially at the lower income deciles where the difference in waiting times are smaller, than at the highest decile.

Regarding the selection equation (specified as the probability of choosing a public hospital) patients with higher income have generally a lower probability of choosing a public hospital. Differences are quantitatively large: compared to the lowest decile, individuals within the highest income decile have a reduced probability of going public by 80%. A negative sign of the Inverse Mill's Ratio shows that people who have a higher propensity of choosing public hospital wait less.

We now move to quantile regression results. Table 3 presents results for the first stage (selection equation) estimation using the Semi Non-Parametric (SNP) estimator as described in Appendix A.[11] The first column presents the results from probit estimates for comparative purposes. The SNP estimates with degree of polynomial of order 3 and 5 (SNP(3) and SNP(5) respectively) are presented in the third and fourth column respectively. We first test if the error term follows a standard Gaussian distribution using the LR test of the probit model against the SNP(3) and SNP(5) models. The assumption of a Gaussian distribution is rejected suggesting that estimation using a probit model will lead to inconsistent estimates of $\boldsymbol{\beta}_{sel}$. We therefore prefer to estimate the coefficients $\boldsymbol{\beta}_{sel}$ using SNP model.[12]

We now compare the coefficients of probit and SNP(5) models. It should be noted that because of different scale normalization and fitted densities, estimated coefficients of the probit model are not directly comparable with those of the SNP models (De Luca, 2008). However, the ratio of estimated coefficients can be compared across models. Here we use the coefficients' ratio with respect to age by normalising the age coefficients to 1. These estimates are presented in the lower half of Table 3.[13] The coefficients on income dummies estimated with the SNP model are very close to those estimated with a probit model.

In summary, the first-stage results for the quantile analysis show that the non-parametric SNP model should be preferred over a simpler probit model (as suggested by the LR test) and

[11] Estimation is done using the '`snp`' command in Stata written by De Luca (2008).

[12] The optimal degree of polynomial of the SNP estimator is chosen using a LR test of order K model versus order K-1 model. We select SNP(5) model as LR tests for K-1 against K reject the null at 1% significance level for K<= 5. The SNP models for higher order are not considered for two main reasons: First, the magnitude of income coefficients from SNP(3) and SNP(5) models are similar suggesting convergence and thus no requirement for higher order models. Second, SNP coefficients which correspond to the coefficients of polynomial terms are all significant at 1% level for model SNP(5) but their magnitudes approach zero for higher orders in SNP(5) model.

[13] The standard error for these ratios is calculated by delta method using "`nlcom`" command in Stata.

therefore over the traditional Heckman selection equation, though the coefficients of income on the probability of choosing public versus private care are very similar (and therefore robust) under the two estimation strategies.

[Table 3 here]

Table 4 reports the quantile regression estimates for three model specifications. Panel A presents the results from quantile regression but does not control for sample selection. Columns 2-6 provide the estimates for different quantiles (10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th}). The results show that compared to patients living in areas with lowest income, patients living in areas in the top three (8-10) income deciles wait 8-12% less across the waiting time distribution. Similarly, patients living in areas with income falling between the 6^{th} and 7^{th} decile wait 3.3-7.7% less across the waiting time distribution. Patients living in areas with income falling between the 2^{nd} and 5^{th} decile wait 3.8-7.7% less at higher quantiles of the waiting time distribution (50^{th}, 75^{th} and 90^{th}) but have no significantly different waiting time than those in the lowest income decile at the lowest quantiles of the waiting time distribution (10^{th} and 25^{th}).

Panel B reports the results for a model specification which controls for sample selection by including the Inverse Mill's Ratio (derived from first stage probit regression) as an additional covariate. The results show that the gradient weakens significantly at lower (10^{th} and 25^{th}) quantiles of the waiting time distribution and it mostly disappears (income is only statistically significant at 10% level for the 8^{th} and 6^{th} decile). At higher quantiles of the waiting time distribution (75^{th} and 90^{th}) the effect of income on waiting times is still negative and significant for individuals whose income falls in deciles 2, 3, 5, 8, 9 and 10. The magnitude of the coefficients is generally reduced, which is in line with the comparison between the results from OLS and the Heckman sample selection model (discussed above).

Panel C reports results for a model specification which controls for sample selection by including a series approximation of bias term (derived from first stage SNP estimation) as additional covariates. The results reveal no evidence of a gradient for 10^{th} and 25^{th} quantile (in line with results in Panel B). The results for 50^{th} quantile reveal a significant negative effect of income on waiting times for six decile dummies, although four of these are significant only at 10% level. The results at the 75^{th} and 90^{th} quantile are similar to those obtained in Panel B though statistical significance is generally smaller.

[Table 4 here]

Table 5 reports the results for four sub-groups of patients: Same-Day Eye surgery, Hip & Knee Surgery, Prostatectomy and Hysterectomy. OLS results show no evidence of a gradient for Hip & Knee Surgery. A gradient is found for Eye Surgery where patients living in areas with income falling between the 7th and 10th decile wait 10-20% less compared to those living in areas with income falling in the lowest decile area. Once selection is controlled for through Heckman selection model, the magnitude of these effects reduce and remain significant at 5% level only for patients whose income falls in the 7th and 8th deciles (the coefficient is about 2% smaller in absolute values). For Prostatectomy patients with income falling in decile 4 and 9 wait respectively 31% and 60% less than patients with lowest income. After controlling for selection the effect on decile 4 ceases to be statistically significant while the effect on decile 9 is significant only at 10% level. For Hysterectomy patients with income in decile 8 wait 21% less both when we control or we do not control for selection (though this is statistically significant only at 10% level).

[Table 5 here]

In summary, the results overall reveal that ignoring sample selection can lead to an overestimation of the income gradient. Part of the effect of income on the time spent on elective surgery waiting lists in Victoria can therefore be attributed to patients with higher income opting for private care. When selection into private hospital care is taken into account, the gradient is reduced.

5. Conclusions

We have investigated the role of sample selection in explaining the gradient between waiting times and socioeconomic status. Using data from the state of Victoria in Australia, we have shown that OLS regression methods suggest that individuals living in areas with higher income wait less in public hospitals compared to individuals living in areas with lower income. More precisely, compared to patients living in areas in the lowest income decile, patients living in areas in the highest income decile wait overall 13% less. Given an average waiting of 89 days, this implies an average reduction of 11 days

Since patients have a choice between public care with a significant waiting time and private care (at negligible waiting time), one possible explanation for this gradient is sample selection: richer patients have a higher probability of choosing private care than poorer ones when waiting times are high. Therefore, the public sector treats poor patients with high and low waiting times but only rich patients with low waiting times. We have controlled for sample selection by estimating Heckman (and related) sample selection models which take into account in the first stage the probability of choosing public versus private care. Our analysis shows that within the state of Victoria in Australia, where about 50% of patients receive private care, sample selection does play a role. Once sample selection is taken into account, the gradient between income and waiting time is reduced and about half of the one estimated by OLS.

There are different possible explanations for the residual gradient between waiting time and socioeconomic status (after taking into account sample selection). Individuals with higher socioeconomic status may engage more actively with the system and exercise pressure when they experience long delays. They may also have better social networks ('know someone') and use them to have priority over other patients, and they may have a lower probability of missing scheduled appointments (which would increase the waiting time). Investigating empirically the different mechanisms which explain the residual gradient (on top of sample selection) is left for future research.

References

Buchinsky, M. (1998). "The dynamics of changes in the female wage distribution in the USA: a quantile regression approach", *Journal of Applied Econometrics,* 13(1): 1-30.

Carlsen, F., O. Kaarboe (2010). "Waiting times and socioeconomic status. Evidence from Norway", Working Papers in Health Economics, 2/10, Health Economics Bergen, University of Bergen, Norway.

Cooper, Z.N., A. McGuire, S. Jones S., J. Le Grand (2009). "Equity, waiting times, and NHS reforms: retrospective study", *British Medical Journal,* 339**,** b3264.

De Luca, G. (2008). "SNP and SML estimation of univariate and bivariate binary-choice models", *Stata Journal,* 8(2): 190-220.

Cullis, P., J.G. Jones, C. Propper (2000). "Waiting and medical treatment: analyses and policies". In A. J. Culyer and J. P. Newhouse (eds.), *North Holland Handbook on Health Economics*, Amsterdam: Elsevier, Chapter 28, 1201-1249.

Duckett, S.J. (2000). "The Australian health care system", Oxford University Press.

Fabbri, D., C. Monfardini (2009). "Rationing the public provision of healthcare in the presence of private supplements: evidence from the Italian NHS", *Journal of Health Economics*, 28(2), 290-304.

Gallant, A.R., D.W. Nychka (1987). "Semi-nonparametric maximum likelihood estimation", *Econometrica,* 55(2): 363-390.

Gravelle, H., L. Siciliani (2008). "Optimal quality, waits and charges in health insurance". *Journal of Health Economics*, 27(3): 663-674.

Heckman, J.J. (1979). "Sample selection bias as a specification error", *Econometrica*, 153-161.

Hoel, M., E.M. Saether (2003). "Public health care with waiting time: the role of supplementary private health care", *Journal of Health Economics*, 22: 599-616.

Johar M., G. Jones, M. Keane, E. Savage, O. Stavrunova, 2010, "Differences in waiting times for elective admissions in NSW public hospitals: A decomposition analysis by non-clinical factors", Working Paper 2010/7, CHERE, University of Technology, Sidney.

Jones, A. (2007). "Applied econometrics for health economists: a practical guide", Radcliffe Publishing.

Koenker, R. (2005). “Quantile regression”, Cambridge University Press.

Iversen, T. (1997). “The effect of a private sector on the waiting time in a national health service.”, *Journal of Health Economics* 16: 381-396.

Laudicella, M., L. Siciliani, R. Cookson (2010). “Waiting Times and Socioeconomic Status: Evidence from England”, University of York, Centre for Health Economics.

Lindsay, C.M., B. Feigenbaum (1984). “Rationing by waiting lists”, *American Economic Review,* 74(3): 404-417.

Marchand, M., F. Schroyen (2005). “Can a mixed health care system be desirable on equity grounds?”, *Scandinavian Journal of Economics*, 107(1): 1-23.

Martin, S., P.C. Smith (1999). “Rationing by waiting lists: an empirical investigation”, *Journal of Public Economics*, 71: 141-164.

Martin, S., P.C. Smith (2003). “Using panel methods to model waiting times for National Health Service surgery”. *Journal of the Royal Statistical Society*, 166/ Part 2: 1-19.

Martin, S., N. Rice, R. Jacobs, R., P.C. Smith (2007). “The Market for Elective Surgery: Joint estimation of supply and demand”, *Journal of Health Economics*, 26(2), 263-285.

Monstad, K., L. B. Engeaeter, B. Espehaug. (2010). “Waiting time socioeconomic status - an individual level analysis”. Working papers in Economics, 11/10, Department of Economics, University of Bergen.

Olivella, P. (2003). “Shifting Public-Health-Sector Waiting Lists to the Private Sector”, *European Journal of Political Economy*, 19: 103-132.

Pagan, A., A. Ullah (1999). “Nonparametric econometrics”, Cambridge University Press.

Siciliani, L., R. Verzulli (2009). “Waiting times and socioeconomic status among elderly Europeans: evidence from SHARE”, *Health Economics*, 18(11): 1295-1306.

Stewart, M.B. (2004). “Semi-nonparametric estimation of extended ordered probit models” *Stata Journal*, 4(1): 27-39.

Tinghög, G., D. Andersson, P. Tinghög, C.H. Lyttkens, 2010, Horizontal Inequality in Rationing by Waiting Lists, mimeo.

Appendix A

First Stage Estimation

The parameters $\boldsymbol{\beta_{sel}}$ in the selection equation (Equation (2)) are estimated using quasi-maximum likelihood (ML) method following a semi non-parametric (SNP) estimator proposed by Gallant and Nychka (1987).[14] However, a maximum likelihood method requires the correct specification of the multivariate probability density function that defines the likelihood, which is unknown.[15] The SNP estimator approximates the density using a Hermite polynomial expansion and uses the approximations to derive pseudo-ML estimator for model parameters. The value of degree of polynomial can be selected by a sequence of likelihood-ratio (LR) tests. The SNP estimator provides consistent estimates if the degree of polynomial increases with sample size.

However, there are two identification issues with this estimator. The intercept coefficient can be absorbed into the unknown distribution of the error term and is not separately identified. Also the slope coefficients can only be identified up to a scale parameter. Thus both location and scale normalizations are required to identify the SNP estimates. Furthermore the scale normalisation must be based on continuous variables with a nonzero coefficient (Pagan and Ullah, 1999). In other words, identification of an SNP estimator requires fixing the value of intercept and having at least one continuous variable in the vector of covariates. Following Melenberg and Soest (1996) we fix the value of intercept to its probit estimate. We also include patient's age as a continuous variable for quantile regression analysis to enable scale normalisation and identification of the SNP estimator.

Second Stage Estimation

Second stage involves approximating the bias term $\mu(Z'\beta_{sel})$ by a series expansion by a polynomial in $Z'\beta_{sel}$, $\widehat{\mu_\tau}(\boldsymbol{Z'\beta_{sel}}) = \hat{\delta}_\tau' P_S(\boldsymbol{Z'}\widehat{\boldsymbol{\beta_{sel}}})$, where $P_S(\boldsymbol{Z'}\widehat{\boldsymbol{\beta_{sel}}}) =$

[14] There are other semi-parametric estimators available, most of which require kernel smoothing, bandwidth selection and trimming to enable convergence. Also these kernel based methods require arbitrary judgement (for example on the choice of bandwidth), are computationally intensive and often have convergence issues with estimation (Stewart, 2004). Therefore we use SNP estimator by Gallant and Nychka (1987) that does not involve kernel smoothing.

[15] It should be noted that the standard approach of a probit model for sample selection (which uses a parametric form of density that defines the likelihood) is not appropriate here as it can bias the estimates of model parameters ($\boldsymbol{\beta_{sel}}$).

$(P_{S1}(\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}), \dots\dots, P_{Sj}(\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}})$ is a polynomial vector of order j in $\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}$. The quantile specific coefficient vector $\hat{\delta}_\tau'$ is obtained from the quantile regression of w_{ij} on h_j , y_{ij}, p_{ij} and $P_S(\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}})$:

$$Q_\tau\big(\ln\big(w_{ij}\big|s_{ij} = 1\big)\big) = h_j(\tau) + \beta_1'(\tau)y_{ij} + \beta_2'(\tau)p_{ij} + \hat{\delta}_\tau' P_S\big(\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}\big) + Q_\tau(u_{1ij}(\tau)|s_{ij} = 1)$$

We use power series expansion of polynomial vector: $P_{Sj}\big(\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}\big) = \lambda\,(\hat{a} + \hat{b}\,\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}\,)^{j-1}$, where λ is the usual inverse Mill's ratio. The normalization of $\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}}$ by $(\hat{a} + \hat{b}\,\boldsymbol{Z}'\widehat{\boldsymbol{\beta}_{sel}})$ is done to make it location and scale equivariant since the semiparametric identification of $\boldsymbol{\beta}_{sel}$ requires scale and local normalization. Following Buchinnsky (1998), the values of $\hat{a}$ and $\hat{b}$ are taken as the constant and slope coefficients respectively from the probit regression of $s_{ij}\ on\ \widehat{\mu}_\tau(\boldsymbol{Z}'\boldsymbol{\beta}_{sel})$. The order j of polynomial is taken as 2 as higher orders were causing severe multicollinearity problems without any significant changes in the value of coefficients. The terms of this polynomial vector are used to calculate the bias term which is used as an additional covariate in Equation (4) to get the quantile regression estimates.

Table 1: Descriptive Statistics

Variables	All categories	Same day Eye Surgery	Hip & Knee Procedures	Hysterectomy	Prostatectomy
Selection Equation (Stage 1) All patients public and private					
Income decile 1	0.080	0.100	0.079	0.090	0.101
Income decile 2	0.144	0.180	0.164	0.150	0.158
Income decile 3	0.086	0.101	0.096	0.091	0.102
Income decile 4	0.108	0.114	0.113	0.098	0.108
Income decile 5	0.100	0.099	0.096	0.096	0.083
Income decile 6	0.117	0.112	0.100	0.127	0.107
Income decile 7	0.100	0.090	0.094	0.096	0.104
Income decile 8	0.092	0.080	0.089	0.085	0.078
Income decile 9	0.106	0.078	0.101	0.100	0.093
Income dec. 10	0.058	0.039	0.060	0.059	0.058
Age <10 years	0.052	0.002	0.000	0.000	0.000
Age 10-20	0.045	0.001	0.000	0.000	0.000
Age 20-30	0.077	0.002	0.009	0.009	0.000
Age 30-40	0.133	0.004	0.135	0.131	0.000
Age 40-50	0.145	0.021	0.444	0.409	0.006
Age 50-60	0.159	0.083	0.214	0.175	0.102
Age 60-70	0.157	0.202	0.109	0.096	0.323
Age 70-80	0.155	0.431	0.066	0.059	0.410
Age >80 years	0.073	0.250	0.019	0.016	0.156
PHI (%)	47.90	37.90	62.10	40.40	46.50
Male	0.434	0.414	0.417	0.000	1.000
Total Patients	219442	16728	9602	3117	3327
Proportion Private Patients	0.60	0.51	0.71	0.45	0.62
Public Hospitals	25	15	22	25	21
Private Hospital	30	19	25	23	23
Waiting Time Regression (Stage 2) waiting list patients treated in public hospitals					
Wait (Days)	88.44	108.42	195.77	81.78	94.77
Income decile 1	0.129	0.154	0.139	0.136	0.156
Income decile 2	0.174	0.180	0.214	0.166	0.169
Income decile 3	0.102	0.113	0.125	0.098	0.114
Income decile 4	0.129	0.125	0.143	0.120	0.143
Income decile 5	0.101	0.100	0.093	0.099	0.089
Income decile 6	0.117	0.106	0.093	0.139	0.108
Income decile 7	0.083	0.070	0.070	0.078	0.089
Income decile 8	0.068	0.062	0.046	0.072	0.055
Income decile 9	0.063	0.058	0.049	0.060	0.043
Income dec. 10	0.026	0.023	0.018	0.025	0.026
Age 10-20	0.052	0.002	0.000	0.000	0.000
Age 20-30	0.090	0.002	0.002	0.014	0.000
Age 30-40	0.124	0.005	0.013	0.166	0.000
Age 40-50	0.143	0.020	0.056	0.453	0.003
Age 50-60	0.135	0.065	0.134	0.182	0.085
Age 60-70	0.147	0.210	0.293	0.102	0.317
Age 70-80	0.158	0.461	0.381	0.067	0.437
Age >80yrs	0.063	0.229	0.118	0.014	0.154
Male	0.439	0.429	0.396	0.000	1.000
Sample Size	88273	8137	2704	1699	1241

Table 2: Results: All Categories

	OLS	Heckman Model	
Variables		Ln(Wait)	Selection Eq.
Income decile 2	-0.057***	-0.029*	-0.352***
Income decile 3	-0.059***	-0.037**	-0.265***
Income decile 4	-0.056***	-0.035**	-0.253***
Income decile 5	-0.063***	-0.030*	-0.387***
Income decile 6	-0.069***	-0.037**	-0.375***
Income decile 7	-0.084***	-0.041**	-0.496***
Income decile 8	-0.112***	-0.068***	-0.466***
Income decile 9	-0.120***	-0.058***	-0.660***
Income decile 10	-0.131***	-0.051*	-0.793***
Private Health Insurance (PHI)	-	-	-3.136***
Inverse Mill's Ratio		-0.220***	
Hospital Fixed effects	√	√	×
DRG Effects	√	√	√
Total Procedures	√	√	√
Primary Diagnosis effects	√	√	√
Age Categories	√	√	√
Gender	√	√	√
PHI	×	×	√
Observations	88273	219442	

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Semi Non-Parametric (SNP) estimates for the selection equation

VARIABLES	Probit	SNP(3)	SNP(5)
Income decile 2	-0.355***	-0.290***	-0.281***
Income decile3	-0.265***	-0.216***	-0.211***
Income decile 4	-0.248***	-0.207***	-0.200***
Income decile 5	-0.401***	-0.332***	-0.321***
Income decile 6	-0.384***	-0.315***	-0.305***
Income decile 7	-0.506***	-0.417***	-0.403***
Income decile 8	-0.470***	-0.396***	-0.381***
Income decile 9	-0.692***	-0.578***	-0.555***
Income decile 10	-0.810***	-0.682***	-0.648***
PHI	-3.136***	-3.159***	-2.991***
SNP Coeff. : 1		-0.405***	-0.595***
2		-0.032***	-0.091***
3		-0.056***	0.148***
4		-	0.011***
5		-	-0.010***
LR test of Probit against SNP model		$\chi2(1)$ = 379.56***	$\chi2(3)$ = 440.66***
LR Test: SNP(3) vs. SNP(5)			$\chi2(1)$ = 75.79***
Ratio of coefficients wrt to age (age coeff. normalised to -1) for direct comparison			
Income decile 2	-1.159***	-1.178***	-1.131***
Income decile3	-0.863***	-0.878***	-0.847***
Income decile 4	-0.826***	-0.821***	-0.804***
Income decile 5	-1.327***	-1.329***	-1.291***
Income decile 6	-1.259***	-1.271***	-1.227***
Income decile 7	-1.666***	-1.678***	-1.622***
Income decile 8	-1.579***	-1.556***	-1.532***
Income decile 9	-2.307***	-2.292***	-2.231***
Income decile 10	-2.722***	-2.684***	-2.606***
PHI	-10.391***	-12.601***	-12.034***

*** p<0.01, ** p<0.05

Table 4: Quantile Regression (All categories)

Variables	q10	q25	q50	q75	q90
A. Without Sample Selection					
Income decile 2	-0.020	-0.017	-0.050***	-0.054***	-0.081***
Income decile3	-0.028	-0.020	-0.050***	-0.066***	-0.070***
Income decile 4	-0.016	-0.040*	-0.052***	-0.042**	-0.039**
Income decile 5	-0.033	-0.034	-0.044**	-0.057***	-0.074***
Income decile 6	-0.069**	-0.068***	-0.052***	-0.037**	-0.037**
Income decile 7	-0.081**	-0.059**	-0.077***	-0.059***	-0.081***
Income decile 8	-0.110***	-0.076***	-0.106***	-0.093***	-0.121***
Income decile 9	-0.102***	-0.083***	-0.093***	-0.106***	-0.137***
Income decile 10	-0.120**	-0.115***	-0.131***	-0.124***	-0.111***
B. With Sample Selection (Standard IMR)					
Income decile 2	-0.000	0.008	-0.025	-0.033**	-0.059***
Income decile 3	-0.013	0.005	-0.033*	-0.050***	-0.055***
Income decile 4	-0.000	-0.020	-0.034*	-0.025	-0.026
Income decile 5	-0.008	0.000	-0.017	-0.033*	-0.049**
Income decile 6	-0.037	-0.036*	-0.027	-0.012	-0.013
Income decile 7	-0.041	-0.018	-0.039*	-0.026	-0.051**
Income decile 8	-0.063*	-0.035	-0.071***	-0.061***	-0.088***
Income decile 9	-0.041	-0.023	-0.036	-0.057***	-0.091***
Income decile 10	-0.057	-0.034	-0.057*	-0.063**	-0.055*
C. With Sample Selection (Series Approximation)					
Income decile 2	0.015	0.020	-0.030	-0.026	-0.059***
Income decile3	0.001	0.013	-0.036*	-0.045**	-0.055**
Income decile 4	0.007	-0.012	-0.037**	-0.019	-0.026
Income decile 5	0.008	0.013	-0.022	-0.024	-0.049**
Income decile 6	-0.022	-0.023	-0.031	-0.003	-0.013
Income decile 7	-0.021	-0.001	-0.044*	-0.015	-0.051*
Income decile 8	-0.046	-0.019	-0.076***	-0.051**	-0.088***
Income decile 9	-0.006	0.000	-0.048*	-0.040	-0.091***
Income decile 10	-0.019	-0.008	-0.070*	-0.044	-0.055

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Results: Sub-groups.

Variables	OLS	Heckman Model Ln(Wait)	Heckman Model Select. Eq.	OLS	Heckman Model Ln(Wait)	Heckman Model Select. Eq.
		Same Day Eye Surgery			**Hip & Knee Surgery**	
Income decile 2	-0.057	-0.031	-0.686***	-0.061	-0.064	-0.255***
Income decile3	-0.009	0.006	-0.395***	0.055	0.054	-0.150
Income decile 4	-0.002	0.014	-0.434***	-0.046	-0.048	-0.181**
Income decile 5	-0.045	-0.025	-0.572***	-0.028	-0.032	-0.454***
Income decile 6	-0.020	-0.002	-0.508***	-0.076	-0.081	-0.400***
Income decile 7	-0.225***	-0.202**	-0.586***	0.088	0.081	-0.552***
Income decile 8	-0.144***	-0.123**	-0.565***	0.010	0.004	-0.666***
Income decile 9	-0.109**	-0.089	-0.509***	-0.024	-0.031	-0.673***
Income decile 10	-0.148*	-0.117	-0.665***	-0.137	-0.148	-1.017***
Inverse Mill's Ratio	N/A	-0.098**		N/A	0.021	
Observations	8,137	16,728		2,704	9,602	

Variables	OLS	Heckman Model Ln(Wait)	Heckman Model Select. Eq.	OLS	Heckmam Model Ln(Wait)	Heckmam Model Select. Eq.
		Prostectomy			**Hysterectomy**	
Income decile 2	-0.011	0.099	-0.727***	-0.090	-0.092	-0.682***
Income decile3	-0.015	0.063	-0.589***	-0.097	-0.098	-0.441**
Income decile 4	-0.309**	-0.260	-0.368***	0.070	0.068	-0.542***
Income decile 5	-0.110	-0.056	-0.419***	0.028	0.027	-0.377**
Income decile 6	-0.061	0.020	-0.614***	-0.073	-0.075	-0.670***
Income decile 7	-0.128	-0.044	-0.542***	-0.026	-0.029	-0.696***
Income decile 8	-0.313	-0.238	-0.570***	-0.207*	-0.210*	-0.733***
Income decile 9	-0.598**	-0.443*	-0.956***	0.066	0.061	-0.987***
Income decile 10	-0.429	-0.299	-0.901***	0.009	0.002	-1.402***
Inverse Mill's Ratio	N/A	-0.29		N/A	0.016	
Observations	1,241	3,327		1,699	3,117	

*** p<0.01, ** p<0.05, * p<0.1