



## Junior doctors' preferences for specialty choice

Peter Sivey<sup>a,\*</sup>, Anthony Scott<sup>a</sup>, Julia Witt<sup>b</sup>, Catherine Joyce<sup>c</sup>, John Humphreys<sup>d</sup>

<sup>a</sup> Melbourne Institute of Applied Economic and Social Research, The University of Melbourne, Australia

<sup>b</sup> Department of Economics, University of Manitoba, Canada

<sup>c</sup> Department of Epidemiology and Preventive Medicine, Monash University, Australia

<sup>d</sup> School of Rural Health, Monash University, Australia

### ARTICLE INFO

#### Article history:

Received 22 December 2011

Received in revised form 11 July 2012

Accepted 12 July 2012

Available online 20 July 2012

#### JEL classification:

I11

J24

#### Keywords:

Discrete choice experiment

Junior doctors

Specialty choice

### ABSTRACT

A number of studies suggest that there is an over-supply of specialists and an under-supply of general practitioners in many developed countries. Previous econometric studies of specialty choice from the US suggest that although income plays a role, other non-pecuniary factors may be important. This paper presents a novel application of a choice experiment to identify the effects of expected future earnings and other attributes on specialty choice. We find the implied marginal wage estimated from our discrete choice model is close to the actual wages of senior specialists, but much higher than those of senior GPs. In a policy simulation we find that increasing GPs' earnings by \$50,000, or increasing opportunities for procedural or academic work can increase the number of junior doctors choosing general practice by between 8 and 13 percentage points. The simulation implies an earnings elasticity of specialty choice of 0.95.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

Economists first observed over 40 years ago that wage differences between primary care physicians and specialists drive specialty choice (Sloan, 1970), which can lead to an excess of specialists. The wage differentials are persistent over time and therefore the market for specialists does not clear (Nicholson, 2002). Wage differences may persist in the long-run because of historically determined and regulated fee or pay bargaining structures (Gagne and Leger, 2005) and barriers to entry (Bhattacharya, 2005). These market imperfections can contribute to an inefficient supply of doctors across specialties leading to sub-optimal health outcomes and high health care costs. In particular, the growing burden of chronic disease suggests that more primary care physicians are needed but there is no market mechanism by which this can be translated into changes in the relative earnings of primary care physicians and specialists. The inefficient distribution of medical practitioners between primary care and other specialties is an ongoing policy issue given the important role that primary care can play in promoting a more efficient and equitable health care system focussed increasingly on chronic disease (Starfield et al., 2005).

Data from the UK show only one quarter of medical graduates aspire to being general practitioners (GPs) although GPs comprise around one half of the senior medical workforce (Lambert et al., 2006). In the US, the number of family medicine positions being chosen by US medical students has halved in the decade to 2009 (Pugno et al., 2010) and a "serious shortage of primary care practitioners is inevitable in the near future" (Bodenheimer and Pham, 2010). In Australia, a retrospective study of cohorts from a large Australian medical school (Joyce and McNeil, 2006) showed a marked decline in the proportion of medical graduates working in general practice, from 52% in the 1980 cohort to 33% in the 1995 cohort. AIHW data shows the number of GPs in Australia grew by 18% between 1998 and 2008 whereas the number of specialists grew by 44% (AIHW, 2010).

The specialty in which doctors practise is determined by the preferences of doctors, and by the demand for different types of doctor as signalled by the number and distribution of specialty training places. Doctors have preferences for expected future earnings and the attributes of each specialty, such as opportunities to use procedural skills, intellectual/academic opportunities, and flexibility of hours.

Although earnings differences between specialties are an issue, other attributes of work/life balance and intrinsic job attributes also influence the choice of specialty. Most studies are from the US and have examined the association between expected future earnings and physicians' actual specialty (Hurley, 1991; Thornton, 2000;

\* Corresponding author at: Level 7, Alan Gilbert Building, Barry St, University of Melbourne, Victoria 3010, Australia. Tel.: +61 3 83442093; fax: +61 3 8344 2111.

E-mail address: [psivey@unimelb.edu.au](mailto:psivey@unimelb.edu.au) (P. Sivey).

Bhattacharya, 2005; McKay, 1990). Other studies have used data on the preference rankings of senior medical students and their actual decisions (Nicholson, 2002; Dorsey et al., 2003). The results have shown that expected future income, working hours (Bhattacharya, 2005; McKay, 1990), 'controllable lifestyle' (Dorsey et al., 2003) and educational debt (Thornton, 2000; Nicholson, 2002) all play important roles. Harris et al. (2005) found that only 16% of junior doctors rate 'financial prospects' of the specialty as important in determining choice whereas more than 50% rated 'hours of work', 'flexible working hours' and 'intellectual content' as important.

The aim of this paper is to provide evidence on the preferences of junior doctors for the different attributes of specialties using a representative sample of Australian postgraduate doctors who have not yet entered a specialty training program. We use our results to simulate policy changes to increase the number of junior doctors choosing general practice as a specialty.

The paper adds to the specialty choice literature in a number of ways. Previous studies have not focussed on the non-pecuniary characteristics that influence doctors' preferences. Though the effect of earnings is usually statistically significant, the size of the elasticity is usually rather small suggesting that a range of other factors play a role (Nicholson and Propper, 2011). It is also often difficult to separate out the attributes of the specialty from the specialty 'label'. For example, surgical specialties often have higher earnings, offer a high level of procedural work, but have less flexible hours, and very little continuity of care. Though expected future earnings can usually be calculated, data on other non-pecuniary attributes are usually unobserved or captured in the constant term in specialty choice models (Nicholson, 2002). It is therefore often difficult to ascertain what it is about a particular specialty that doctors value, and how these factors are traded-off. This is a particular issue for policy makers where it has not always been possible to alter relative earnings as a policy response. Alternative policy responses could be found by altering non-pecuniary attributes. This study therefore uses a stated preference discrete choice experiment to examine both pecuniary and non-pecuniary factors that influence specialty preferences. To focus on the attributes of each specialty, rather than use specialty labels (e.g. surgery or general practice) for each choice, we use an unlabelled experimental design (i.e. choice between specialty A and B).

Second, Nicholson (2002) observes that earlier studies which used data after a specialty was chosen, assume that the market clears and that observed specialties represent those most preferred. The assumption that the market clears is essential to be able to interpret results as the preferences of doctors. With competition for specialty training places combined with markets not clearing, doctors may end up practising in their second or third most preferred specialty. These earlier studies therefore do not directly examine preferences of doctors as the market does not clear. To address this issue, Nicholson (2002) examines stated preference rankings rather than actual specialty destinations and estimates an income elasticity higher than previous studies. We use a discrete choice experiment focussing on stated preferences using a sample of junior doctors who have not yet chosen their specialty.

Finally, studies that use data on specialty attributes and actual specialty choices may be subject to endogeneity bias in estimating the effect of income and other factors on junior doctors' specialty choices. Income may be correlated with unobservable doctor or specialty attributes (e.g. ability of the doctor or prestige of the specialty) which could bias the estimate of the income coefficient. Gagne and Leger (2005) attempt to solve this problem using Canadian data where changes over time in government-regulated physician fee levels, which are exogenous to hours worked or productivity, are used as a proxy for income. However their study provides little information on other (non time-varying)

determinants of specialty choice. Another example of potential endogeneity is the inclusion of educational debt in some models of specialty choice. In Thornton (2000) educational debt is positively associated with choosing general practice (a lower remunerated specialty). This may be indicative of educational debt being correlated with unobservable factors associated with the choice of general practice.

Identification of the effect of each attribute on specialty choice can be achieved in a DCE in two ways. First, through the use of an efficient experimental design where attributes can be independently varied and standard errors minimised. The experimental choice context ensures that choices are presented to respondents exogenously. Second, since a DCE presents respondents with multiple choices for each respondent, then respondents' unobserved heterogeneity (e.g. ability) can be accounted for in the econometric analysis. We use a generalised multinomial logit model that allows for preference heterogeneity and also allows for scale heterogeneity, a form of heteroskedasticity (Fiebig et al., 2010, 2011).

We also contribute to the DCE literature in health economics where there are still relatively few applications of the technique in analysing healthcare labour markets (Lagarde and Blaauw, 2009). Most papers have studied job choice, especially for GPs (e.g. Scott, 2001; Ubach et al., 2003) or rural/urban location issues (e.g. Kolstad, 2011; Gunther et al., 2011). This paper is the first discrete choice experiment analysing the specialty choice of junior doctors, a group whose preferences are arguably the most important for the future of the health workforce.

## 2. Methods

### 2.1. Data—the MABEL survey

The "Medicine in Australia: Balancing Employment and Life (MABEL)" study investigates workforce participation patterns and their determinants using a longitudinal survey of Australian doctors. All Australian doctors undertaking clinical work ( $n = 54,750$ ) were invited to participate in the first wave in 2008. Data are collected by paper or optional online version of a questionnaire, with content tailored to four sub-groups of clinicians: general practitioners, specialists, specialists in training, and hospital non-specialists. The survey methods are discussed in detail in Joyce et al. (2010). The survey included discrete choice experiments tailored to the different types of doctors answering the survey. In this paper we report on the results from the discrete choice experiment administered to hospital non-specialists. We define our sample as junior doctors who have completed their basic medical degree but have not yet enrolled in a specialty training program or as a GP registrar. Our sample includes first year interns, and those in their first three years as a hospital (or resident) medical officer (HMO/RMO). In 2008 there were 5284 interns and HMOs (or RMOs) in Australia (AIHW, 2010).

### 2.2. Development of attributes and levels

The choice experiment consists of presenting survey respondents with a series of hypothetical choices between two alternative specialties. Each alternative specialty is defined by a list of attributes.

The attributes used in the questionnaire were developed through consulting previous international literature on healthcare labour markets and specialty choice, as well as literature on which attributes might be important from previous research in Australia (AMWAC, 2005; Harris et al., 2005). We conducted a pre-pilot where a group of junior doctors gave feedback about the range

**Table 1**  
Specialty attributes and levels.

Specialty attribute	Levels
Expected average annual earnings	\$150,000 \$200,000 <sup>a</sup> \$250,000
Change in total hours worked	10% increase The same <sup>a</sup> 10% decrease
Control over hours	Low Medium <sup>a</sup> High
On-call arrangements	1 in 2 1 in 4 <sup>a</sup> 1 in 4 – infrequently called out 1 in 10
Opportunities for procedural work	None Some <sup>a</sup> Enough
Academic/research opportunities	Poor Average <sup>a</sup> Excellent
Continuity of care	Rarely see patients more than once Sometimes see patients more than once <sup>a</sup> Regularly see patients more than once

<sup>a</sup> Reference category.

and wording of the attributes and levels. In general junior doctors were familiar with the attributes used in the experiment as they all relate to aspects of their working life they have experienced. We did not undertake extensive qualitative research to determine the attributes (Coast et al., 2012) as previous studies from overseas and Australia (e.g. Harris et al., 2005) and policy issues guided our choice of key attributes, which were then confirmed in the pre-pilot. Although our choice of attributes was informed by a prior literature which uses ex-post questions about the determinants of specialty choice, our study uses only the ex-ante responses from doctors who have yet to enter a specialty program.

Table 1 lists the attributes and levels included in the experiment. The first four attributes can be regarded as ‘work-life’ attributes in the sense that they affect the doctor’s out-of-work life, including consumption and leisure time. The final three attributes can be regarded as ‘intrinsic’ specialty attributes, as they relate to characteristics of the doctor’s experiences in the workplace.

The earnings attribute allows us to measure marginal willingness-to-pay (or willingness-to-accept) for changes in all other attributes and especially the valuation of reducing working hours. We can also use the earnings attribute to simulate the earning elasticity of specialty choice (Sloan, 1970; Nicholson, 2002). Working hours are a key attribute of any job and are important from a policy perspective because there is a trend towards working fewer hours in Australia and the USA (McKay, 1990; Scott, 2006; Staiger et al., 2010). We also included on-call arrangements since this is a key issue for most doctors. Reasonable on-call, weekend and after-hours duties have been found to be important job characteristics, with lower out-of-hours workloads being preferred (Scott, 2001; Ubach et al., 2003).

All of the attributes represent characteristics of alternative specialties in the future. For the earnings attribute we must take into account that junior doctors may expect large increases in their annual earnings over their career. For this reason we specify

‘expected average annual earnings’ to make the choice plausible for respondents. There are no nationally representative data on doctors’ earnings in Australia. Instead we chose the level of the earnings attribute by checking through classifieds for physician jobs across a variety of specialties on various association and hiring agencies’ websites and using these salary ranges to calculate our bounds. We allow for a fairly wide range of future earnings to include a range of specialties, including general practice, we chose: 150,000, 200,000 and 250,000 Australian dollars.

For the change in hours worked attribute, we chose 10% decrease, no change, 10% increase. This was chosen to match approximately the variation in average total hours worked per week across different specialties (AIHW, 2009).

We defined four levels for the on-call attribute: 1 in 10, frequently called out; 1 in 4, infrequently called out; 1 in 4, frequently called out; 1 in 2, frequently called out. With these attribute levels, we wanted to reflect two dimensions of the on-call attribute (1) the official on-call arrangements (e.g. 1 in 4), and (2) the frequency of call outs in each on-call period. The terminology reflects the proportion of days a doctor will be on-call (e.g. 1 in 10 indicates 1 in every 10 days). We obtained this on-call terminology from classified job advertisements.

In addition to hours worked and on-call, the final work/life attribute is control over hours. Harris et al. (2005), using Australian survey data, found that flexible work hours were important in choosing a specialty, and Dorsey et al. (2003) found that controllable lifestyle was increasingly important in explaining physician specialty preference in the US. For control over hours, we have the following three levels: high, medium, low. We wanted this range to be broad to account for the different settings in which control over hours will vary considerably. For example, specialists in their own private practice probably have more control over their hours than specialists in public hospitals.

The final three attributes relate only to qualitative aspects of the time spent at work. The first of these attributes is academic opportunities. AMWAC (2005) found that one of the three most influential factors in choice of specialty was the intellectual content of the specialty. Intellectual content has also been found to be an important determinant of specialty choice in a qualitative Canadian study (Horn et al., 2008). We include three levels to measure the extent of preference for academic opportunities: excellent, average, poor.

The second intrinsic attribute is continuity of care. Stokes et al. (2005) conducted a three-country study of the importance of continuity of care among family physicians/general practitioners, and found that all place a high value on being able to provide continuity of care to their patients. Continuity of care is a defining attribute of primary care, but less so for other specialties. For continuity of care, we have the following three levels: regularly see patients more than once; sometimes see patients more than once, rarely see patients more than once. These levels are likely to capture two components: (1) differences between specialties, for example, GPs having a higher rate of continuity than, for example, anaesthesiology, and (2) workplace arrangements, where, for example, working in a hospital is likely to be associated with less continuity, while working in a private practice is likely to be associated with more continuity.

Finally, AMWAC (2005) found that opportunities for procedural work were among the most influential determinants of choice of specialty in Australia. Recent research has shown procedure-based specialties are increasingly being chosen by Canadian junior doctors (Horn et al., 2008). Procedural work provides the opportunity to use technical skills, which some doctors value. For opportunities for procedural work, we have the following three levels: Enough, some, none. The levels include the spectrum ranging from GPs,

many of whom do only minor procedural work, to surgical specialists, who do mostly procedural work.

### 2.3. Experimental design

Our experimental design is based on the logit model. The model arises from a job-characteristics approach to choice of specialty. This approach has previously been used to analyse job choice for GPs (e.g. Scott, 2001; Lagarde and Blaauw, 2009) and is based on a random utility model. We specify an indirect utility function where utility for respondent  $i$  from choosing specialty  $j$  in choice situation  $t$  is a linear combination of attributes of specialty  $j$  and an idiosyncratic error term  $\varepsilon_{ijt}$ .

$$U_{ijt} = X_{ijt}\beta + \varepsilon_{ijt} \quad (1)$$

where  $X_{ijt}$  is a vector containing the attributes of alternative specialties. The model is based on doctors' hypothetical choices between two alternative specialties, specialty 1 and specialty 2 ( $j = 1, 2$ ). The choice is modelled by assuming  $\varepsilon_{ijt}$  is type 1 extreme value (Gumbel) distributed giving the probability that doctor  $i$  chooses specialty 1 according to the logistic distribution function:

$$\Pr(U_{i1t} - U_{i2t} > 0) = \frac{\exp(X_{i1t}\beta)}{\exp(X_{i1t}\beta) + \exp(X_{i2t}\beta)} \quad (2)$$

Given the model specified by (1) and (2), we use an approach to developing D-efficient experimental designs as described by Zwerina et al. (1996) and Carlsson and Martinsson (2003). This involves using prior information about coefficient values to minimise the expected standard errors and therefore improve the statistical efficiency of the estimated choice models. The design process was completed twice: first for a pilot survey, and then for the main survey. The pilot of the DCE included a sample of 62 junior doctors. Logit models estimated with the data from the pilot survey were then used to provide prior values of the betas for the main survey design.

There are seven attributes, six with three levels, and one with four levels giving a full factorial of  $3^6 \times 4^1 = 2916$  possible choices. For the pilot survey and then for the main survey, we generated a fractional factorial of 36 binary choice sets containing 72 alternatives. The 36 choice sets were blocked into four versions and each doctor completed 9 unlabelled choices each.

Our choice of an unlabelled experimental design may be criticized as less realistic than a labelled design (Kruijschaar et al., 2009). However, studies in the health economics (De Bekker-Grob et al., 2010) and environmental economics (Blamey et al., 2000) literature have demonstrated that labelled designs can reduce the attention respondents give to attribute levels, with up to 24% of respondents choosing based on choice labels alone. An unlabelled design encourages respondents to focus on the attributes in the experiment. We think an unlabelled design is especially suitable for our context as labelled choices could cause the results to be dominated by junior doctors' unobserved preconceptions about certain specialties.

Another aspect of our design is that we use a 'forced choice' experiment where respondents must choose one of the two options (the hypothetical specialties). Some authors have previously argued 'forced choices' are inappropriate (Ryan and Skatun, 2004; King et al., 2007) and that DCEs should include an 'opt-out' or 'neither' option. Opt-out options increase realism in a DCE when there is a possible bias for the status-quo due to unobservable attributes. For the population we investigate in this paper, we argue a forced choice is appropriate because junior doctors in Australia are in a temporary role as an intern or HMO and are all likely to choose a specialty training program in which to enrol in the near future. Even if they remain as a medical officer this can be regarded

as a form of specialty choice, which is different from their current role. For this reason, we do not include an 'opt-out' choice in this experiment.

An efficient design was generated for the pilot by minimizing the D-error, a function of the covariance matrix ( $\Sigma$ ) of the  $\beta$ 's in the logit model.

$$D\text{-error} = |\Sigma|^{1/K} \quad (3)$$

where  $K$  is the number of  $\beta$ s to be estimated. D-errors are lower with better priors since  $\Sigma$  depends on the parameter values, and thus the better the estimates, the better the approximation of the covariance matrix that is used to generate the efficient design. For the pilot survey, we set our priors to zero, because we had no prior information.

We used a SAS program (Zwerina et al., 1996) to search for the best design using a modified Fedorov candidate set search algorithm. We used a full factorial candidate set, since this did not increase running time. The best design among those generated was found by comparing D-errors and choosing the one with the lowest D-error that also had the most "sensible" choice pairs. We tried to avoid designs with too many attribute combinations that respondents may not have found very realistic. For example, regular continuity of care may be associated with GP's, but not with procedural specialties, and so designs in which these were frequently paired to describe one specialty were discarded in favour of designs that paired these attribute levels less frequently. The main survey DCE had a D-error of 0.25.

Respondents were randomly allocated to one of the four blocks of choice sets in the questionnaire.

The experiment is designed to capture 'main effects' only. While allowing for interaction effects between attributes is desirable (Lancsar and Louviere, 2008) they should only be included where there are clear hypotheses about interactions. Furthermore the relatively large number of attributes and levels in our experiment makes identifying these effects unfeasible.

An example of the choice experiment is given in Fig. 1.

### 2.4. Econometric estimation

We first estimate a logit model using the data collected in the choice experiment. The earnings and hours attributes are coded as continuous variables. All of the other attributes are dummy-coded with the 'middle' category of the attribute values as the reference category (see Table 1). Eq. (2) defines a logit model assuming  $\varepsilon_{i1t}$  and  $\varepsilon_{i2t}$  are independently and identically distributed for all respondents  $i$  and across all choice situations  $t$ .

Recent literature (e.g. Hall et al., 2006; Hole, 2008) has used the mixed logit model to allow for unobserved heterogeneity in the coefficients of the model. Where the coefficient vector  $\beta$  is assumed to follow a distribution with density  $f(\beta)$ , the choice probability given by Eq. (2) for the logit model is extended by integrating the choice probability over the density of  $\beta$ :

$$\Pr(U_{i1t} - U_{i2t} > 0) = \int \frac{\exp(X_{i1t}\beta_i)}{\exp(X_{i1t}\beta_i) + \exp(X_{i2t}\beta_i)} f(\beta_i|\Theta) d\beta \quad (4)$$

The mixed logit can be further extended to the generalised multinomial logit model (Fiebig et al., 2010). This model extends the mixed logit model to allow for scale heterogeneity. Scale heterogeneity is a form of heteroskedasticity where the variance of the error term varies across respondents, but is the same between choices for the same respondent. This form of heteroskedasticity implies a different 'scale' of the coefficient distributions for each respondent. Scale heterogeneity is particularly a concern in choice experiments (Flynn et al., 2010) where respondents may vary in



Please read the following:

- Imagine that you are choosing a specialty in which to work, and you have the choice between two specialties, A and B.
- Everything about the two specialties is the same, except for the characteristics shown in the tables below.
- Even if you would not choose either specialty, we would like you to state which you think is better.

9. Which specialty (A or B) would you prefer?

	Specialty A	Specialty B
Expected average annual earnings	\$150,000	\$200,000
Change in total hours worked	10% decrease	No change
On-call arrangements	1 in 4, infrequently called out	1 in 10, frequently called out
Control over hours	Medium	Low
Academic/Research opportunities	Average	Poor
Continuity of care	Sometimes see patients more than once	Rarely see patients more than once
Opportunities for procedural work	None	None

☐ Prefer Specialty A ☐ Prefer Specialty B

Fig. 1. Example of the DCE preamble and question.

how ‘certain’ they are in their responses to a DCE. One can also think of scale as representing how some respondents answer the DCE more ‘seriously’ than others.

Respondents who have a high degree of certainty about their responses, and take the DCE exercise seriously will have a low variance of the error term and a correspondingly high scaling of the coefficient vector. Respondents who are uncertain of their responses or who take the DCE exercise less seriously will have a high variance of the error term and a relatively low scaling of the coefficient vector. This model has recently been developed in the marketing literature (Fiebig et al., 2010) and has already been applied in health economics (Fiebig et al., 2011).

We estimate a GMNL model where the coefficient vector  $\beta_i$  in the choice probability given in Eq. (4) for individual  $i$  can be characterized as follows:

$$\beta_i = \sigma_i(\bar{\beta} + \tilde{\beta}_i) \quad (5)$$

where the scalar  $\sigma_i = \exp(\bar{\sigma} + \tau v_i)$  is the scaling factor,  $v_i \sim N(0, 1)$  and  $\tilde{\beta}_i \sim N(0, \theta^2)$ . The vectors  $\bar{\beta}$  and  $\theta$  are the means and standard deviations of normally distributed random coefficients (as in the mixed logit model). The parameter  $\bar{\sigma}$  is set at  $\tau^2/2$  and so the only extra parameter to be estimated<sup>1</sup> is  $\tau$ , which measures the degree of scale heterogeneity. Overall the parameters to be estimated are given by the vector  $\Theta = (\bar{\beta}, \theta, \tau)$ , where  $\bar{\beta}$  is a vector of means of the coefficients,  $\theta$  is a vector of the standard deviations of the coefficients and  $\tau$  is a scalar measuring scale heterogeneity.

We use 1000 Halton draws in the NLOGIT software package (Greene and Hensher, 2010) for the estimations. Our main aim with the GMNL model is to test which attributes have substantial preference and scale heterogeneity as represented by the standard deviations of the estimated coefficient distributions and the  $\tau$  parameter.

Median marginal willingness to pay (MWTP) values are calculated for all attributes except the earnings attributes for both the logit and GMNL models. MWTP for an attribute is the ratio of the coefficient estimate for that attribute and the coefficient estimate for the earnings attribute.

## 2.5. Interaction terms

We use a GMNL model with interaction terms<sup>2</sup> between the specialty attributes and selected individual characteristics to test two specific hypotheses arising from the literature.

**Hypothesis 1.** Junior doctors with high levels of educational debt will value future earnings more highly.

This hypothesis is informed by the prior literature from the US which tests whether educational debt is an important factor in specialty choice (Thornton, 2000; Nicholson, 2002). These papers hypothesise (1) that junior doctors with more debt may choose better-remunerated specialties (therefore have a stronger preference for earnings) to be able to repay their debt more easily, or conversely (2) that junior doctors with more debt may choose specialties with shorter required training periods. The empirical findings of these papers are mixed, with no clear confirmation of either hypothesis. The present study allows us to test Hypothesis 1 in Australia, where university education is more heavily subsidized than the US, but substantial educational debts are still present. The Higher Education Contribution Scheme (HECS) is the main government-run student loan and repayment system in Australia. This scheme provides student loans to cover tuition with interest rates indexed to the rate of inflation and repayment only required when income is above certain minimum thresholds. The generosity of the student loan scheme should lessen the financial incentives associated with educational debt in Australia compared to the US.

To test the hypothesis we interact the level of educational debt with the earnings attribute in the GMNL model. The coefficient on this interaction will represent the degree to which educational debt affects the respondent's preferences for future earnings. The relevant question for educational debt in MABEL is: “What is the total level of financial debt that you currently have as a result of your medical education and training? (give dollar amount; include HECS debt, other debt associated with training and living expenses)”.

**Hypothesis 2.** Female doctors and doctors with children will value flexibility of hours worked more highly.

Harris et al. (2005) find that “Factors of particular importance to women, compared with men, were “Appraisal of domestic circumstances” (odds ratio, 1.9), “Hours of work” (OR, 1.8) and “Opportunity to work flexible hours” (OR, 2.6)” in a retrospective study of specialty choice. Flexible hours may be more important for

<sup>1</sup> Note we estimate the GMNL-II version of the model given in Fiebig et al. (2010), where the scale parameter affects the mean and standard deviation of the coefficient distributions equally. In the full unrestricted version of the GMNL model the  $\gamma$  parameter appeared to be poorly identified so we set it equal to 0.

<sup>2</sup> Our ‘interaction term’ characteristics enter the GMNL model as determinants of the mean of the coefficient distribution for the relevant attribute.

women as they often take a majority role in childcare. As MABEL also has detailed information on domestic circumstances, including children in the family, we also test for the effect of having children in the family (for male and female doctors).

To test this hypothesis we interact two dummy variables: “Female” and “Children” (=1 if the doctor reports having any children) with the “Change in hours”, “Control over hours” and “On-call” attributes.

## 2.6. Predicted probability analysis and policy simulation

To learn more about the policy implications of the results we conduct a predicted probability analysis (Lancsar and Louviere, 2008) and policy simulation. This involves using the estimated coefficients to simulate the model (calculating predicted probabilities for each choice) for a plausible real life choice between alternative specialties. Secondly, we simulate the model after unilateral changes in attributes for one of the alternative specialties. These changes in attributes can represent government policy changes.

Due to the shortage of GPs in Australia (AMWAC, 2005), and primary care physicians internationally (Bodenheimer et al., 2007) we simulate the choice of “General Practitioner” versus the choice of “Specialist”.

We simulate the two predicted probabilities given below:

$$Pr(GP|X_{GP}, X_{SPEC}, \hat{\Theta}) = \int \frac{\exp(X_{GP}\beta_i)}{\exp(X_{GP}\beta_i) + \exp(X_{SPEC}\beta_i)} f(\beta_i|\hat{\Theta}) d\beta \quad (6)$$

$$Pr(SPEC|X_{SPEC}, X_{GP}, \hat{\Theta}) = \int \frac{\exp(X_{SPEC}\beta_i)}{\exp(X_{GP}\beta_i) + \exp(X_{SPEC}\beta_i)} f(\beta_i|\hat{\Theta}) d\beta \quad (7)$$

These predictions use the estimated parameters from the models from Eqs. (4) and (5) where  $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}, \hat{\tau})$  are the estimated coefficient means, standard deviations and scale parameter. We also specify two groups of attribute levels  $X_{GP}$  and  $X_{SPEC}$  which aim to represent a ‘typical’ choice between a specialist training program and general practice. The attribute levels are informed by evidence from the MABEL survey itself and from the literature.

In order to increase the realism of our policy simulation we calibrate our model to ensure the predicted choice probabilities match the actual proportion of junior doctors currently choosing general practice (vs specialist training). Otherwise, our model predictions may diverge from these true proportions due to unobserved attributes of general practice or specialties that are not captured in our model. We follow the procedure implemented by Fiebig et al. (2011) and suggested by Train (2009, Section 7.2) to introduce a constant term,  $\alpha^r$ , into the coefficient vector for the choice of ‘GP’ that captures these unobserved attributes. Given we have revealed-preference data on the actual proportion of junior doctors choosing “GP”, which we denote by  $Pr(\widehat{GP})$ , we can use an iterative procedure of simulating probabilities with alternative values of  $\alpha^r$  to solve the following equation for the constant  $\alpha^r$ :

$$Pr(GP|X_{GP}, X_{SPEC}, \hat{\Theta}, \alpha^r) = \int \frac{\exp(\alpha^r + X_{GP}\beta_i)}{\exp(\alpha^r + X_{GP}\beta_i) + \exp(X_{SPEC}\beta_i)} \times f(\beta_i|\hat{\Theta}) d\beta = Pr(\widehat{GP}) \quad (8)$$

In other words, we find the value of  $\alpha^r$  for which the simulated probability given by Eq. (8) matches the true values of the proportion of junior doctors choosing to be GPs.

**Table 2**  
Descriptive statistics.

	Obs	Mean	S.D.
Age	532	28.782	4.643
Female	536	0.615	0.487
Children (>0)	536	0.124	0.330
Educational debt (0,000 AUD)	444	2.772	3.535
Income (gross AUD per year)	286	73,161	38,391
Hours/week	476	50.046	11.613
Position			
Intern	536	0.180	0.385
HMO yr 1	536	0.306	0.461
HMO yr 2	536	0.299	0.458
HMO yr 3	536	0.214	0.411

The simulation then proceeds by considering changes to the attribute levels  $X_{GP}$  that represent potential policy changes that would affect the proportion of junior doctors choosing to train as GPs.

## 3. Results

The MABEL response rate for “Hospital non-specialists” is 16.45% from a sampling frame of 8820 giving 1451 respondents. After excluding pilot respondents and career medical officers<sup>3</sup> we have 536 respondents (interns and HMOs/RMOs) who answer at least one DCE question. The estimation sample is 4808 observations from 536 junior doctors. Joyce et al. (2010) find the survey is broadly representative of the population and we also examine representativeness in our sample descriptive statistics given in Table 2. The mean age is 28.8 years old, 61.5% of the sample is female and the mean working hours is 50.0 h a week. These values are close to the most comprehensive estimate of the population of interns and HMOs/RMOs (AIHW, 2010) for age (28.7 years) and hours worked (49.0 h). However it appears we over-represent female junior doctors (52.7% female in the AIHW data). We test for the effect of gender on the model coefficients in the interactions model.

Table 3 reports estimates for the logit and GMNL models. For both models we present the Bayesian Information Criterion ( $BIC = -2 \log L + k \times \ln(n)$ ) where  $\log L$  is the log-likelihood,  $k$  is the number of parameters estimated and  $n$  is the number of observations.

In the logit model, the estimated coefficients are statistically significant at 1% for all attributes apart from “Continuity of care – Regularly”. The estimated coefficients all have the expected sign, and for the attributes with three dummy-coded levels, utility is estimated to be monotonically increasing in the attribute. Doctors prefer lower hours of work, high control over hours, low on-call, excellent academic opportunities, and high levels of procedural work.

The GMNL model coefficient means are all approximately twice the size of the coefficient point-estimates in the logit model, which is has been found in other studies using this model (Greene and Hensher, 2010). All but one of the coefficient distributions have substantial and statistically significant standard deviations. This suggests that there is substantial preference heterogeneity over the attributes. The lower BIC in the GMNL indicates this model is preferred in terms of model ‘fit’ and the statistically significant estimate of the ‘scale parameter’ ( $\tau$ ) indicates there is scale heterogeneity amongst respondents.

<sup>3</sup> ‘Career medical officers’ are doctors who have chosen to remain as ‘hospital non-specialists’ rather than enter a specialty training program.

**Table 3**

Logit and GMNL model results.

Attribute	(1) Logit		(2) GMNL		
	Coeff. (S.E.)	MWTP	Coeff. (S.E.)	S.D. (S.E.)	MWTP
Earnings (\$0,000)	0.096*** (0.006)		0.184*** (0.016)	0.014*** (0.002)	
Change in hours (%)	−0.040*** (0.003)	−4216	−0.075*** (0.008)	0.032** (0.012)	−4053
Control hours – high	0.206*** (0.057)	21,529	0.369*** (0.110)	0.695*** (0.254)	20,059
Control hours – low	−0.525*** (0.053)	−54,752	−0.965*** (0.116)	0.614*** (0.204)	−52,529
On-call – 1 in 10	0.713*** (0.070)	74,401	1.339*** (0.143)	0.341 (0.416)	72,857
On-call – 1 in 2	−1.097*** (0.071)	−114,530	−2.088*** (0.192)	1.374*** (0.244)	−113,592
On-call – 1 in 4 infrequent	0.602*** (0.071)	62,869	1.151*** (0.146)	1.037*** (0.169)	62,628
Academic – excellent	0.240*** (0.049)	25,016	0.439*** (0.095)	0.906*** (0.160)	23,862
Academic – poor	−0.333*** (0.053)	−34,722	−0.708*** (0.110)	0.935*** (0.168)	−38,498
Continuity – rarely	−0.312*** (0.058)	−32,593	−0.601*** (0.113)	0.843*** (0.170)	−32,694
Continuity – regularly	0.058 (0.062)	6024	0.129 (0.110)	0.943*** (0.152)	7004
Procedural – enough	0.209*** (0.050)	21,814	0.381*** (0.098)	0.834*** (0.162)	20,732
Procedural – none	−0.601*** (0.059)	−62,671	−1.138*** (0.131)	1.053*** (0.200)	−61,941
Tau (scale parameter)			0.278*** (0.040)		
log L	−2617.29		−2501.63		
BIC	5345		5232		
Obs	4808		4808		

Notes. Model (1), logit; Model (2), generalised multinomial logit (GMNL). Model (2) assumes the normal distribution for the coefficients of all attributes and allows for scale heterogeneity in the coefficient distributions. Marginal willingness to pay (MWTP) values are the ratio of coefficient estimates for each attribute, and the coefficient estimate for earnings. For the GMNL, given all coefficients have a normal distribution, this ratio is the median of the unconditional distribution of MWTP. The reference category is Control over hours – medium, On-call – 1 in 4, Academic – average, Continuity – sometimes, Procedural – some.

\* Statistical significance at 10%.

\*\* Statistical significance at 5%.

\*\*\* Statistical significance at 1%.

The “On-call – 1 in 10” attribute coefficient has a small standard deviation which is not statistically significant. This indicates most doctors have similar preferences over this attribute, in this case it has a positive effect on utility; doctors prefer less on-call time. “Change in hours (%)” also has a relatively small standard deviation (<40% of the coefficient mean) indicating relatively little variation in preferences over working fewer hours.

Marginal willingness-to-pay (MWTP) values, the marginal rate of substitution between each attribute coefficient and the earnings coefficient, are also calculated in Table 3. We can interpret these values as the sum of money (in terms of future annual income) a doctor would give up in order to gain a unit increase in the attribute. A negative value represents an attribute with a negative effect on utility and can be interpreted as the amount of expected future income the doctor would be willing to accept (in terms of annual income) to compensate for a unit increase in the attribute.

The two models produce very similar values<sup>4</sup> for most of the attributes. We concentrate on the MWTP for the GMNL model. In general, all the specialty attributes have large but plausible monetary values according to our estimates. An exception is the “On-call” attribute which has very high valuations, outside the range of the earnings attribute (in the case of “On-call – 1 in 2”). A consistent

result is that the higher (‘better’) level of each attribute is valued less over the medium level than the medium level is valued over the lowest level. This finding is consistent with diminishing returns to attributes and suggests that losses are valued more than gains.

Both work-life and intrinsic specialty attributes have substantial valuations. We estimate that doctors are willing to accept a \$53,000 decrease in expected annual earnings to have “Control over hours – Medium” rather than “Control over hours – Low”. A slightly higher valuation (\$62,000) is given to having some procedural work compared to none. Attributes with lower valuations are “Academic opportunities” and “Continuity of care”. Both of these attributes have the lower levels valued at \$33,000–\$39,000.

### 3.1. Predicted probability analysis and policy simulation

We proceed to a simulation exercise predicting choice probabilities using the coefficients estimated in model (2) (the GMNL model). For the baseline simulation, we use MABEL data to inform the attribute levels for the two alternative choices. Where we quote figures in the following text, they are the raw means of the corresponding variable presented in Table 4.

For the earnings and hours attributes we have direct measures in MABEL. In terms of gross earnings, specialists earn on average \$334,937 and GPs earn on average \$183,067, a difference of \$151,870. For the simulation we set the difference between the two alternatives to be \$150,000.

<sup>4</sup> For the GMNL model we present the median of the unconditional distribution of MWTP.

**Table 4**

Mean values of MABEL variables used to inform policy simulations.

Variable	GPs	Specialists
Earnings (gross annual \$AUD)	183,067	334,937
Hours/week	38.8	45.4
Hours of work: very/moderately dissatisfied	0.17	0.28
It is difficult to take time off when I want to: strongly agree	0.39	0.43
The hours I work are unpredictable: agree	0.21	0.44
On-call ratio	7.66	5.87

**Table 5**

Model simulations from Model (2).

Attribute	GP	Specialist
Earnings	\$180,000	\$330,000
Change in hours	–15%	0%
Control over hours	Medium	Low
On-call	1 in 10	1 in 4
Academic	Poor	Average
Continuity	Regularly	Sometimes
Procedural work	None	Enough
	% choose GP	% choose specialist
Base case	33.9	66.1
Calibrated base case	39.9	60.1
Increase procedural work to “Some”	53.0	47.0
Increase earnings by 50,000 to \$230,000	50.4	49.6
Increase academic opps to “Average”	47.8	52.2

Notes. The bottom half of the table presents predicted probabilities (expressed as percentages) from Model (2) in Table 3 when attribute levels are as described in the top half of the table. ‘Calibrated base case’ (and all three scenarios that follow) introduces a constant of the value 0.5178 into the utility function of the ‘GP’ choice.

For hours worked per week (not including on-call), specialists work 45.4 h and GPs 38.8 h, a difference of 6.6 h. For the simulation we choose a 15% difference, as 15% of the specialists’ mean working hours is 6.6 h.

Evidence from three MABEL variables shows that, compared to GPs, specialists are more likely to be dissatisfied with their hours of work (28% vs 17%), agree that they cannot take time off when they want to (43% vs 39%), and have unpredictable hours (44% vs 21%). Using this evidence in the simulation, we set “Control over hours” to be “Medium” for GPs and “Low” for specialists.

As GPs see the same patient for multiple medical problems, and due to the continuous/ongoing nature of primary care for chronic disease and family planning, GPs generally provide more continuity of care to their patients than specialists. In the simulation, we set “Continuity of care” to be “Sometimes” for specialists and “Regularly” for GPs. For “Academic opportunities” there is evidence in Australia (Joyce et al., 2009) that more specialists than GPs are involved in research (15% vs 1%), so we set this attribute to “Average” for specialists and “Poor” for GPs.

Using the relevant MABEL question we find that, on average, GPs are on-call 1 in 7.7 and specialists 1 in 5.9. To relate these to the ratios used in the DCE attribute, for the simulation we choose On-call to be “1 in 10” for GPs<sup>5</sup> and “1 in 4” for specialists.

Table 5 presents the attribute levels for the simulation and the predicted probabilities for four simulations of the model. The uncalibrated base case predicts that 33.9% of junior doctors will choose the “GP” option and 66.1% will choose “Specialist”.

To complete the simulation exercise, we follow the procedure detailed in Section 2.6, to calibrate the model to actual market

shares. Using data from the Australian Medical Training Review Panel (2010) we find that in 2009 there were 2352 junior doctors commencing the second year of postgraduate training in Australia and 938 first-year GP trainee positions available, giving a proportion of 39.9%. We then calibrate the model by introducing a constant of the value 0.5178 to the ‘GP’ alternative which increases the predicted probability of the GP alternative to 39.9%.

The final three rows of the table show how unilateral changes in three selected attributes for the “GP” alternative affect the predicted probabilities.

We can see how increasing procedural work from “None” to “Some” has the largest effect on choice probabilities, increasing the number of doctors choosing general practice from 39.9% to 53.0%. Increasing earnings by \$50,000 has a slightly smaller effect, increasing the proportion choosing general practice to 50.4%. Giving GPs “Average” instead of “Poor” academic opportunities increases the proportion by only 7.9 percentage points to 47.8%. The ranking of the effects of these different attributes corresponds to their ranking according to marginal willingness-to-pay (see Table 3, \$61,941, \$50,000 and \$38,498).

The simulated effect of the change in earnings on the choice probability can also be expressed in terms of an elasticity. The earnings elasticity for the simulated choice of ‘GP’ can be expressed as  $(\partial \Pr(\text{GP})/\partial \text{earnings})(\text{earnings}/\Pr(\text{GP}))$ . Table 5 shows that a change in earnings of \$50,000 gives a change in probability of choosing GP of 10.5 percentage points, implying an earnings elasticity of specialty choice of  $(10.5/50,000) \times (180,000/39.9) = 0.95$ .

Table 6 presents results for a GMNL model with interaction terms informed by the two hypotheses in Section 2.5. Table 6 presents results where “Educational debt”, “Female” and “Children” are used as interactions.

**Hypothesis 1.** Doctors with high levels of educational debt will value future earnings more highly.

The coefficient on the interaction between educational debt and earnings is statistically significant at the 5% level, providing some evidence of an effect. The earnings coefficient is 0.166 and the educational debt interaction is 0.007 so a 100% reduction in debt from \$27,000 (the sample mean) to zero decreases the marginal utility of earnings by 0.019 or 12%.

**Hypothesis 2.** Female doctors and doctors with children will value flexibility of hours worked more highly.

The effects of the “Female” and “Children” dummy variable interactions are all statistically insignificant at conventional levels. *F*-tests for joint significance of both groups of interactions also fail to reach statistical significance.

#### 4. Discussion

This paper has focussed on the pecuniary and non-pecuniary factors influencing the preferences of junior doctors for specialty choice. The results show a range of work-life and intrinsic job attributes influence choice of specialty for junior doctors. Doctors would be prepared to sacrifice substantial proportions of their expected annual income, 20–25% based on an annual income of \$200,000,<sup>6</sup> for improvements in control over working hours and opportunities to do procedural work. The most highly valued attribute is time spent on-call, where avoiding being on call every other day had a valuation nearly twice as high as any other attribute (approximately 50% of annual income). Previous research

<sup>5</sup> Note this ratio is unlikely to be realistic for rural GPs who often have to be on-call much more often.

<sup>6</sup> \$200,000 is the middle value of annual earnings in the choice experiment.



**Table 6**

GMNL model with educational debt, gender and children interactions.

Attributes	Coeff. (S.E.)	S.D. (S.E.)	Attributes	Coeff. (S.E.)	S.D. (S.E.)
Earnings (\$0,000)	0.166*** (0.019)	0.123*** (0.022)	On-call – 1 in 2	–1.853*** (0.270)	1.463*** (0.242)
int: ed debt (\$0,000)	0.007** (0.003)		int: female	–0.472 (0.290)	
Change in hours (%)	–0.071*** (0.011)	0.028 (0.018)	int: children	0.200 (0.403)	
int: female	–0.008 (0.011)		On-call – 1 in 4 infreq	0.976*** (0.218)	0.915*** (0.178)
int: children	0.010 (0.016)		int: female	0.307 (0.248)	
Control hours – high	0.204 (0.183)	0.757*** (0.216)	int: children	–0.184 (0.373)	
int: female	0.257 (0.222)		Academic – excellent	0.404*** (0.102)	1.015*** (0.169)
int: children	0.157 (0.324)		Academic – poor	–0.605*** (0.111)	0.737*** (0.182)
Control hours – low	–0.879*** (0.176)	0.504** (0.202)	Continuity – rarely	–0.560*** (0.119)	0.759*** (0.192)
int: female	0.013 (0.199)		Continuity – regularly	0.174 (0.125)	1.100*** (0.163)
int: children	0.123 (0.288)		Procedural – enough	0.327*** (0.103)	0.775*** (0.167)
On-call – 1 in 10	1.326*** (0.218)	0.591** (0.269)	Procedural – none	–1.222*** (0.144)	1.145*** (0.182)
int: female	–0.001 (0.237)				
int: children	–0.253 (0.336)				
Tau (scale parameter)	0.000 (0.0380)				
Obs			4020		
log L			–2065.90		
BIC			4464		

Notes. Generalised multinomial logit (GMNL) assuming normal distribution for all main effect attribute coefficients and allows for scale heterogeneity in coefficient distributions. Interaction variables are preceded by 'int:' and are determinants of the mean of the coefficient distributions for the attribute above. Reference category is Control over hours – medium, On-call – 1 in 4, Academic – average, Continuity – sometimes, Procedural – some.

\* Statistical significance at 10%.

\*\* Statistical significance at 5%.

\*\*\* Statistical significance at 1%.

has shown GPs in the UK also place very high values on avoiding time spent on-call (Scott, 2001).

In interpreting these values we must recognize how the choice of levels for the earnings attribute could influence estimated valuations (Skjoldborg and Gyrd-Hansen, 2003). As we have chosen a relatively wide range for earnings, we may expect this to provide lower valuations for earnings, or equivalently high monetary valuations of other attributes, including time spent on-call. However, our choice of levels was informed by evidence relating to expected future earnings, and hence the relatively wide range is appropriate.

Our estimates however still suggest a substantial effect of earnings on specialty choice. Our estimated earnings elasticity of specialty choice (choice of GP in our simulation) at 0.95 is lower than the 1.42 estimated by Nicholson (2002) but higher than other previous research (Thornton and Esposto, 2003 estimate an elasticity of 0.58). This elasticity estimate suggests changes in GPs earnings will result in approximately equal proportional changes in the number of junior doctors choosing to become a GP. For example, increasing GPs earnings by 10% would increase the number of GPs by 9.5%.

The “Change in hours” attribute suggests doctors will trade-off hours worked per week at a \$4053 change in annual salary for a 1% change in hours. We can convert this monetary value into a hypothetical marginal wage rate, or the additional wage required to encourage a doctor to work an extra hour. Using the figure for

average hours worked from Table 2 we can see 1% of the average weekly hours worked is 0.50 h. Converting the annual salary into a weekly amount we have \$4,053/52 = \$77.95 for a change of 0.50 h per week, so we have a hypothetical marginal wage rate of \$77.95/0.50 = \$155.89 per hour. We can interpret this as the wage at which the average junior doctor would be prepared to work an extra hour when choosing between specialties.

As the MABEL data include information on earnings and hours for specialists and GPs, we can compare the wage rate implied by the DCE with the actual hourly wages of specialists and GPs. Cheng et al. (in press) show that the average hourly wage for a GP is \$89.30 and for a specialist \$145.50. The finding that the marginal wage implied by our coefficient estimates is relatively close to the actual average wage of specialists in the MABEL survey (the implied marginal wage rate is only 7% higher) adds validity to our results. It also provides evidence of a gap between the earnings expectations of junior doctors and the average hourly wage of GPs (\$89.30). The earnings gap is a common explanation for oversupply of specialists and undersupply of GPs (Bodenheimer et al., 2007).

The closest previous study to ours is Nicholson (2002), who use a conditional logit model to examine the preferences of US medical students. We improve on this by explicitly including a range of non-pecuniary attributes, by avoiding potential endogeneity through the use of a discrete choice experiment, and through analysis using a generalised multinomial logit model that accounts for preference and scale heterogeneity.

Our simulations can be used to see how changes in attributes of GP working conditions may influence the future GP workforce. Our calibrated simulation predicts a baseline probability of choosing 'GP' of 39.9% that rises by 13.1 percentage points with an increase in procedural work or 10.5 percentage points (26%) for a \$50,000 (28%) increase in earnings (implying an earnings elasticity of 0.95). Increasing academic opportunities to 'Average' increases the probability by 7.9 percentage points. Based on the 2009 level of 938 first-year GP training positions, an increase of 7.9, 10.5, or 13.1 percentage points, from a baseline of 39.9%, as suggested by our simulation would translate to 186, 247 or 308 extra doctors choosing to enroll as GP registrars every year.

The policy simulation exercise could be extended to evaluate many different alternative bundles of attributes, representing individual specialties. While out of the scope of this paper, this could be a subject for future research.

Previous revealed preference studies from the US have tested whether educational debt influences specialty choice through a preference for higher earnings (Thornton, 2000; Nicholson, 2002). Our results give some limited support to this hypothesis in a different context where the financial incentives associated with educational debt are not as high given government subsidies for higher education in Australia. In contrast, we have not been able to corroborate previous research from Australia suggesting female doctors or doctors with children have a higher valuation of flexible working hours and a preference for shorter working hours (Harris et al., 2005). This result is surprising because of our strong prior expectation that female doctors would on average have stronger preferences for flexible working. For example, the growing dominance of female doctors in the general practice workforce is often explained by the flexibility general practitioners are afforded. Only 12% of doctors in our sample have children, and so family reasons may be less important for junior doctors at this early stage in their career.

This study uses a stated-preference approach which allows us to exogenously vary attributes of specialties when presenting respondents with choice scenarios. This approach also allows us to analyse the effects of attributes that may be unobserved in revealed-preference data. Future research could include using revealed preference data on junior doctors' choices together with a calibrated choice experiment on a sample of these doctors.

One potential criticism of this study is that our survey has a relatively low response rate (Joyce et al., 2010). However, our sample has been shown to be broadly representative of the population based on age and hours worked and it has been shown that response bias does not always diminish with higher response rates (Schoenman et al., 2003).

## 5. Conclusion

This paper has policy implications for addressing the shortage of GPs (AMWAC, 2005). Firstly, our findings underline the importance of earnings in determining specialty choice. The finding that our estimates of doctors' valuation of their future desired wages is much higher than GPs' current wages, and close to the actual wages of specialists, suggests that many junior doctors are expecting higher incomes than are available in general practice. Addressing the GP/specialist earnings differential could be an effective policy solution, though this is an area that is difficult to address. In the US the introduction of the resource-based relative value scale in Medicare did attempt to alter relative prices paid, but evidence suggests that physicians compensated through altering the volume of care they provide (McKay, 1990; Maxwell et al., 2007). Price regulation alone may not therefore be effective at changing relative

earnings. The introduction of the Quality and Outcomes Framework for GPs in the UK, a large performance pay scheme, led to increases in GP incomes of around 25%. It is not yet clear what impact this has had on the proportion of junior doctors choosing general practice as a career. In Australia, GPs and private specialists are free to charge patients what the market will bear, with a fixed subsidy from Medicare. Altering the subsidy for GP visits may alter incomes, but would also increase demand if the gap between the fee charged and subsidy fell. Australia is also experimenting with a limited pay for performance scheme, but these schemes do not always involve new money, rather a re-allocation of payments, and so may not increase earnings (Scott et al., 2010).

In addition to changing relative GP incomes, our results predict that increasing procedural work or academic opportunities for GPs could have a similar effect as increasing earnings by \$50,000, and increase the number of doctors choosing to train as GPs by between 186 and 308 per year. The Primary Health Care Research Evaluation and Development strategy was put in place by the Australian Government in 2000 to try and increase academic activity within general practice. The importance of academic departments of general practice has also been noted in the USA (Newton and DuBard, 2006).

The reality that much of general practice is non-procedural (Baron, 2009) sits uncomfortably with our finding that young doctors prefer more procedural work. Doctors' preference for procedural work may be related to their training and skills or to fee-for-service payment models that generally remunerate procedures better than non-procedural activity. In recent years in Australia, many GPs have been forced to reduce traditional areas of procedural practice, such as obstetrics. It may be that strategies to better support these activities (for example, changed medical indemnity insurance arrangements) would enhance the attractiveness of general practice and reduce referrals to hospitals. Australian GPs in rural areas often undertake more procedural work than their metropolitan colleagues, and this feature could be used to enhance recruitment into these regions, which have long suffered from workforce shortages.

## Acknowledgments

This work was supported by a National Health and Medical Research Council Health Services Research Grant (454799) and the Commonwealth Department of Health and Ageing. The views in this paper are those of the authors alone. We thank the doctors who gave their valuable time to participate in MABEL, and the other members of the MABEL team for data cleaning and comments on drafts of this paper: Terence Cheng, Sung-Hee Jeon, Danny Hills, Guyonne Kalb, Daniel Kuehnle, Anne Leahy, Matthew McGrail, Michelle McIsaac, Stefanie Schurer, Durga Shrestha and Wenda Yan. We are also thankful for comments from participants at the Australian Labour Econometrics Workshop 2011. The study was approved by the University of Melbourne Faculty of Business and Economics Human Ethics Advisory Group (Ref. 0709559) and the Monash University Standing Committee on Ethics in Research Involving Humans (Ref. CF07/1102–2007000291).

## References

- Australian Institute of Health and Welfare, 2009. National Health Labour Force Series. Medical Labour Force 2007. AIHW, Canberra.
- Australian Institute of Health and Welfare, 2010. Medical labour force 2008. Detailed tables—hospital non-specialists. <<http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442468557>> (accessed 01.09.11).
- Australian Medical Workforce Advisory Committee, 2005. Career decision making by postgraduate doctors. AMWAC Report 2005.3. <<http://www.ahwo.gov.au/documents/Publications/2005/Career%20decision%20making>

- %20by%20postgraduate%20doctors%20-%20%20Key%20findings.pdf> (accessed 31.05.11).
- Baron, R.J., 2009. The chasm between intention and achievement in primary care. *Journal of the American Medical Association* 301 (18), 1922–1924.
- Bhattacharya, J., 2005. Specialty selection and lifetime returns to specialization within medicine. *Journal of Human Resources* 40 (1), 115–143.
- Blamey, R.K., Bennett, J.W., Louviere, J.J., Morrison, M.D., Rolfe, J., 2000. A test of policy labels in environmental choice modelling studies. *Ecological Economics* 32 (2), 269–286.
- Bodenheimer, T., Berenson, R.A., Rudolf, P., 2007. The primary care-specialty income gap: why it matters. *Annals of Internal Medicine* 146, 301–306.
- Bodenheimer, T., Pham, H.H., 2010. Primary care: current problems and proposed solutions. *Health Affairs* 29 (5), 799–805.
- Carlsson, F., Martinsson, P., 2003. Design techniques for stated preference methods in health economics. *Health Economics* 12, 281–294.
- Cheng, T.C., Scott, A., Jeon, S.H., Kalb, G., Humphreys, J., Joyce, C. What factors influence the earnings of GPs and medical specialists in Australia? Evidence from the Medicine in Australia: balancing employment and life survey. *Health Economics*, <http://dx.doi.org/10.1002/hec.1791>, in press.
- Coast, J., Al-Janabi, H., Sutton, E.J., Horrocks, S.A., Vosper, A.J., Swancutt, D., Flynn, T.N., 2012. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics* 21 (6), 730–741.
- De Bekker-Grob, E.W., Hol, L., Donkers, B., Van Dam, L., Habbema, J.D.F., Leerdam, M.E., Kulpers, E.J., Essink-Bot, M.L., Steyerberg, E.W., 2010. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value in Health* 13 (2), 315–323.
- Dorsey, E.R., Jarjoura, D., Rutecki, G.W., 2003. Influence of controllable lifestyle on recent trends in specialty choice by US medical students. *Journal of the American Medical Association* 290 (9), 1173–1178.
- Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science* 29, 393–421.
- Fiebig, D.G., Knox, S.A., Viney, R., Haas, M., Street, D., 2011. Preferences for new and existing contraceptive products forthcoming. *Health Economics* 20 (S1), 35–52.
- Flynn, T.N., Louviere, J.J., Peters, T.J., Coast, J., 2010. Using discrete choice experiments to understand preferences for quality of life. *Variance-scale heterogeneity matters*. *Social Science and Medicine* 70 (12), 1957–1965.
- Gagne, R., Leger, P.T., 2005. Determinants of physicians' decisions to specialize. *Health Economics* 14, 721–735.
- Greene, W.H., Hensher, D.A., 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* 37 (3), 413–428.
- Gunther, O.H., Kurstein, B., Riedel-Heller, S.G., Konig, H., 2011. The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. *Health Services Research* 45 (1), 212–229.
- Hall, J., Fiebig, D.G., King, M.T., Hossain, I., Louviere, J.J., 2006. What influences participation in genetic carrier testing? Results from a discrete choice experiment. *Journal of Health Economics* 25 (3), 520–537.
- Harris, M.G., Gavel, P.H., Young, J.R., 2005. Factors influencing the choice of specialty of Australian medical graduates. *Medical Journal of Australia* 183 (6), 295–300.
- Hole, A.R., 2008. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics* 27, 1078–1094.
- Horn, L., Tzanetos, K., Thorpe, K., Straus, S.E., 2008. Factors associated with the subspecialty choices of internal medicine residents in Canada. *BMC Medical Education* 8, 37.
- Hurley, J., 1991. Physicians' choices of specialty, location, and mode: a reexamination within an interdependent decision framework. *Journal of Human Resources* 26 (1), 47–71.
- Joyce, C.M., McNeil, J.J., 2006. Fewer medical graduates are choosing general practice: a comparison of four cohorts, 1980–1995. *Medical Journal of Australia* 185, 102–104.
- Joyce, C.M., Piterman, L., Wesselingh, S.L., 2009. The widening gap between clinical, teaching and research work. *Medical Journal of Australia* 191, 169–172.
- Joyce, C.M., Scott, A., Jeon, S., Humphreys, J., Kalb, G., Witt, J., Leahy, A., 2010. The 'Medicine in Australia: Balancing Employment and Life (MABEL)' longitudinal survey—protocol and baseline data for a prospective cohort study of Australian doctors' workforce participation. *BMC Health Services Research* 10, 50.
- King, M.T., Hall, J., Lancsar, E., Fiebig, D., Hossain, I., Louviere, J., Reddel, H., Jenkins, C., 2007. Patient preferences for managing asthma: results from a discrete choice experiment. *Health Economics* 16 (7), 703–717.
- Kolstad, J.R., 2011. How to make rural jobs more attractive to health workers. Findings from a discrete choice experiment in Tanzania. *Health Economics* 20 (2), 127–252.
- Kruijshaar, M.E., Essink-Bot, M.L., Donkers, B., Looman, C.W.N., Siersema, P.D., Steyerberg, E.W., 2009. A labeled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. *BMC Medical Research Methodology* 9, 31.
- Lagarde, M., Blaauw, D., 2009. A review of the application and contribution of discrete choice experiments to inform human resources policy interventions. *Human Resources in Health* 7, 62.
- Lambert, T.W., Goldacre, M.J., Turner, G., 2006. Career choices of United Kingdom medical graduates of 2002: questionnaire survey. *Medical Education* 40, 514–521.
- Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics* 26 (8), 661–677.
- Maxwell, S., Zuckerman, S., Berenson, R.A., 2007. Use of physicians' services under medicare's resource-based payment. *New England Journal of Medicine* 356, 1853–1861.
- McKay, N.L., 1990. Economic determinants of specialty choice by medical residents. *Journal of Health Economics* 9 (3), 335–357.
- Medical Training Review Panel, 2010. Thirteenth Report. April 2010. Commonwealth of Australia, Canberra.
- Newton, W.P., DuBard, C.A., 2006. Shaping the future of academic health centers: the potential contribution of departments of family medicine. *Annals of Family Medicine* 4 (Suppl. 1), S2–S11.
- Nicholson, S., 2002. Physician specialty choice under uncertainty. *Journal of Labor Economics* 20 (4), 816–847.
- Nicholson, S., Propper, C., 2011. Medical workforce. In: Pauly, M., McGuire, T.G., Pita-Barros, P. (Eds.), *Handbook of Health Economics*. Elsevier, Amsterdam, Ch. 14.
- Pugno, P.A., McGaha, A.L., Schmittling, G.T., DeVilbiss Bieck, A.D., Crosley, P.W., Ostergaard, D.J., 2010. Results of the 2010 National Resident Matching Program: family medicine. *Family Medicine* 42, 8.
- Ryan, M., Skatun, D., 2004. Modelling non-demanders in choice experiments. *Health Economics* 13, 397–402.
- Schoenman, J.A., Berk, M.L., Feldman, J.J., Singer, A., 2003. Impact of differential response rates on the quality of data collected in the CTS physician survey. *Evaluation and the Health Professions* 26, 23–42.
- Scott, A., 2001. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics* 20, 329–347.
- Scott, A., 2006. The productivity of the health workforce. *Australian Economic Review* 39 (3), 312–317.
- Scott, A., Naccarella, L., Furler, J., Young, D., Sivey, P., Ait Ouakrim, D., Willenberg, L., 2010. Using Financial Incentives to Improve the Quality of Primary Care in Australia. Report to Australian Primary Health Care Research Institute. Australian National University, Canberra.
- Sloan, F., 1970. Lifetime earnings and physicians choice of specialty. *Industrial and Labour Relations Review* 24 (1), 47–56.
- Skjoldborg, U.S., Gyrd-Hansen, D., 2003. Conjoint analysis. The cost variable: an Achilles' heel? *Health Economics* 12, 479–491.
- Staiger, D.O., Auerbach, D.I., Buerhaus, P.I., 2010. Trends in the work hours of physicians in the United States. *Journal of the American Medical Association* 303 (8), 747–753.
- Starfield, B., Shi, L., Macinko, J., 2005. Contribution of primary care to health systems and health. *Milbank Quarterly* 83 (3), 457–502.
- Stokes, T., Tarrant, C., Mainous 3rd, A.G., Schers, H., Freeman, G., Baker, R., 2005. Continuity of care: is the personal doctor still important? A survey of general practitioners and family physicians in England and Wales, the United States, and The Netherlands. *Annals of Family Medicine* 3 (4), 353–359.
- Thornton, J., 2000. Physician choice of medical specialty: do economic incentives matter? *Applied Economics* 32 (11), 1419–1428.
- Thornton, J., Esposto, F., 2003. How important are economic factors in choice of medical specialty? *Health Economics* 12, 67–73.
- Train, K., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
- Ubach, C., Scott, A., French, F., Awramenko, M., Needham, G., 2003. What do hospital consultants value about their jobs? A discrete choice experiment. *British Medical Journal* 326, 1432–1437.
- Zwerina, K., Huber, J., Kuhfeld, W., 1996. A general method for constructing efficient choice designs. Working Paper, Fuqua School of Business, Duke University.