CHERE
CENTRE FOR HEALTH ECONOMICS
RESEARCH AND EVALUATION
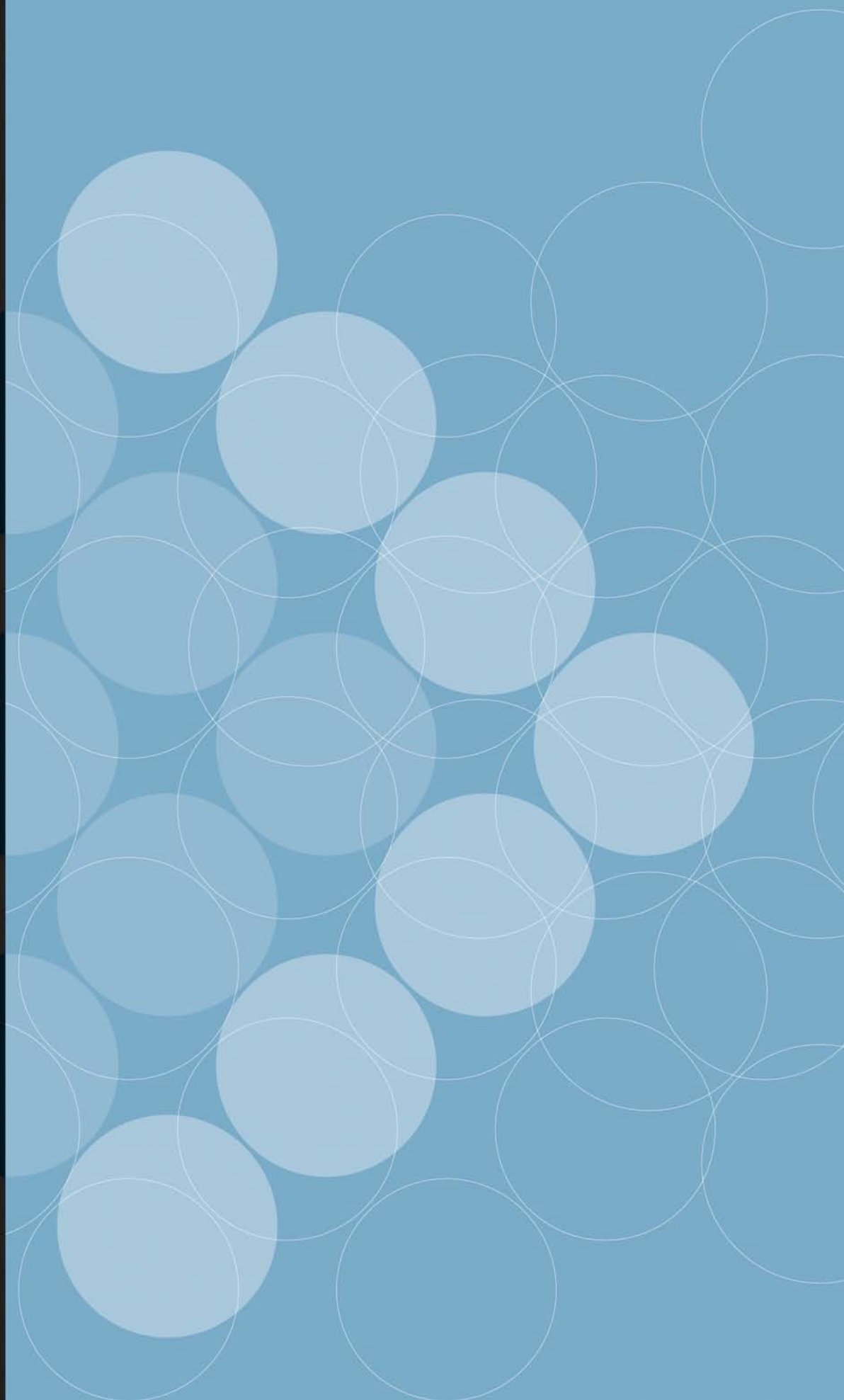
Job Preferences of Students and New Graduates in Nursing

Working Paper 2011/02

July 2011

A report by the Centre for Health Economics Research and Evaluation

CHERE
CENTRE FOR HEALTH ECONOMICS
RESEARCH AND EVALUATION

About CHERE

CHERE is an independent research unit affiliated with the University of Technology, Sydney. It has been established since 1991, and in that time has developed a strong reputation for excellence in research and teaching in health economics and public health and for providing timely and high quality policy advice and support. Its research program is policy-relevant and concerned with issues at the forefront of the sub-discipline.

CHERE has extensive experience in evaluating health services and programs, and in assessing the effectiveness of policy initiatives. The Centre provides policy support to all levels of the health care system, through both formal and informal involvement in working parties, committees, and by undertaking commissioned projects. For further details on our work, see www.chere.uts.edu.au.

Project team

Denise Doiron[1]
Jane Hall[2]
Patricia Kenny[2]
Deborah J. Street[3]

1. School of Economics, Australian School of Business, University of New South Wales
2. CHERE, University of Technology, Sydney
3. School of Mathematical Sciences, University of Technology, Sydney

Corresponding author:
Email: d.doiron@unsw.edu.au Phone: 612 9385 3734 Fax: 612 9313 6337.

Contact details

Denise Doiron, School of Economics, Australian School of Business,
University of New South Wales, Australia, 2052.
Email: d.doiron@unsw.edu.au Phone: 612 9385 3734 Fax: 612 9313 6337

Abstract

This paper investigates the preferences of student and newly graduated nurses for pecuniary and non-pecuniary aspects of nursing jobs. It is the first study applying DCE methods to a developed country nursing workforce. It is also the first to focus on the transition through university training and into work; this is particularly important as junior nurses have the lowest retention levels in the profession. We sample 526 individuals from nursing programs in two Australian universities. Flexible and newly developed models combining heteroskedasticity with unobserved heterogeneity in scale and preference weights are estimated. Overall, salary remains the most important feature in increasing the probability that a job will be selected as best. Supportive management/staff and quality of care follow as the most important attributes from a list of 11 non-pecuniary job characteristics. Newly graduated nurses rank supportive management/staff above salary increases, implying that a supportive workplace is important for the transition from university to the workforce. We find substantial preference heterogeneity and some attributes, such as the opportunity for clinical rotations, are found to be attractive to some nurses while seen as negative by others. Nursing retention could be improved by designing different employment packages to appeal to these different tastes.

1 Introduction

Nurses comprise the largest professional group in the health care workforces of most developed countries; adequate numbers are needed not only to ensure health services delivery, but a workforce with appropriate qualifications and skill levels are important for the quality of health care (Aiken et al., 2002; Heinz, 2004; Needleman and Hassmiller, 2009). Many countries are facing shortages in supply, which are expected to increase in the future, with population ageing and the growing incidence of disability (Oulton, 2006). The evidence from Europe, North America, and Australia suggests that the nursing workforce is aging, with many nurses likely to retire within the next decade. At the same time, the expansion of nursing roles in primary care, chronic disease management and preventive services is an important component of reforms aimed at improving the efficiency and affordability of health systems (Productivity Commission, 2005; Rother and Lavizzo-Mourey, 2009).

Workforce attrition for reasons other than retirement is also a contributor to nursing shortages with pre-retirement age nurses leaving to change careers, and females with dependents more likely to leave the workforce (Nooney et al., 2010). Although relatively low pay rates make nursing less attractive compared to other occupations, it seems that wage elasticities are generally low (Shields, 2004). If increasing pay levels generate only modest impact on increasing workforce participation, then policy makers can turn to increasing supply through attracting more students to training. This has been the policy approach by various governments. However, attrition rates among young and newly registered nurses are high (Barron and West, 2005; Doiron et al., 2008; Naude and McCabe, 2005; Fochsen et al., 2006) and there is evidence that the transition from student to registered nurse can be particularly stressful (Casey et al., 2004). To date, very little is known about the causal mechanisms behind the poor retention rates immediately following graduation from nursing programs.

There is a growing body of evidence that non-pecuniary factors are significant in improving nursing retention (Shields and Ward, 2001): for example, part-time or full time work (Di Tommaso et al., 2009; Zeytinoglu et al., 2011), hours worked (Di Tommaso et al., 2009) opportunities for further training (Frijters et al., 2007); stress and high workloads (Zeytinoglu et al., 2006), supportive work environments (Zeytinoglu et al., 2011), having management responsibilities (Frijters et al., 2007), and sector/type of facility (Di Tommaso et al., 2009). There is also evidence of heterogeneity in retention across nurses (Frijters et al., 2007; Cunich and Whelan, 2010) making further explorations of how individual characteristics and circumstances affect supply, an important consideration for the development of suitable policies (Antonazzo et al., 2003). Beyond the suggestion that working conditions should accommodate the needs for women with young families, there has been little investigation of this (Doiron et al., 2008; Cunich and Whelan, 2010).

The available data sets for studying the nursing labour force, primarily general household surveys or registration data, do not contain sufficiently rich information to allow for detailed study of this range of factors. While surveys enable the researcher to collect more detailed individual data, they are often limited by the range of jobs and job characteristics currently in place, particularly where pay rates and other conditions are set centrally. Stated preference techniques have become an increasingly popular approach to overcome the lack of revealed preference data. Perhaps surprisingly, given the widespread popularity of discrete choice methods in health economics, there are few applications to job preferences; the survey by Lagarde and Blauuw (2009) identifies nine such studies. A handful of these include nurses, all of them are set in a developing country context

and investigate nurses' willingness to take jobs in rural locations (Blaauw et al., 2010; Mangham and Hanson, 2008; Penn-Kekana et al., 2005). All studies demonstrate the importance of wages and non-pecuniary benefits, including the opportunity for further education and training, adequate equipment and infrastructure.

Alongside the growing use of DCEs, there has been increasing attention to the appropriateness of the methods, both for survey design and for the analysis of data. The standard use of multinomial logit models (MNL) has been overtaken by the use of the mixed logit models (MXL) to better account for heterogeneity in preferences across individuals (Keane and Wasi, 2009). While the importance of preference heterogeneity can be considered well established, recent contributions also point to the importance of scale heterogeneity; that is, differences across individuals in utility variance, often interpreted as an individual's uncertainty over preferences. The generalized multinomial logit (GMNL) has been developed to address both scale and preference heterogeneity (Fiebig et al., 2010; Keane and Wasi, 2009). Indeed Fiebig et al. (2010) conclude that scale heterogeneity is relatively more important where decisions are complex, and identify health decisions as a case in point; on the other hand, Greene and Hensher (2010) argue that emphasis on scale heterogeneity over preference heterogeneity may be misguided.

This study focuses on nurses' preferences over jobs, a significant and continuing policy issue, and it investigates preferences for both pecuniary and non-pecuniary aspects of nursing jobs. As mentioned earlier, existing studies of this kind are based on developing countries and it is possible that tradeoffs between monetary and non-monetary job characteristics differ in developed economies. We use data from DCEs involving Australian nursing students and new graduates.

A second novel aspect of the study is the focus on nurses through their training and transition from education to the workforce. As already noted, nurses in these years are especially vulnerable to attrition. The experiences in the early years of training and working as a nurse may well influence motivation and preferences over different job attributes. It is likely that young students choose nurse training without experience on the wards and so have little idea of what it feels like to work as a nurse. Although Registered Nurses (RNs) in Australia receive their education at universities, their education includes classroom learning, simulated experiences in laboratory tutorials and clinical placements in hospitals where they observe and practice nursing work in a structured and supervised way. Their job preferences may be influenced as they experience what nurses actually do. In the analysis, we distinguish job preferences of nursing students according to the year in the program and graduation status.

The study also contributes to the literature by implementing state-of-the art econometric models, some of which are new developments. We compare results from standard MNL and MXL models to the newly developed GMNL model. In addition, we exploit the use of best-worst choice information and estimate rank-ordered and heteroskedastic versions of the MNL, MXL and GMNL models. Best-worst judgments are argued to be both easier tasks for respondents and a means of obtaining more information (Flynn et al., 2007; Vermeulen et al., 2010) compared to the standard approach that asks for the preferred choice only. In this study, respondents are presented with a choice of three job options each described in terms of attribute-levels, and asked to select the best and worst options.[1] A large number of attributes are used reflecting the complexity of actual nursing jobs.

---

[1] Best-worst choices have been structured in different ways and the approach used here is sometimes referred to as best-worst alternative. A different method is the 'best-worst attribute level' where respondents are presented with one option, described in terms of attributes/levels and asked to select the best feature and the worst feature (Flynn et al., 2007).

The rest of the paper is set as follows: Section 2 describes the DCE and the development of attributes; Section 3 gives a fuller description of the sample and data collection; Section 4 reports the model specification and tests; Section 5 discusses the results; Section 6 reports the results for preferences by time in program; Section 7 concludes.

2 The Choice Experiment

Theoretically larger choice sets (scenarios) give more information than do smaller ones but of course considering a large number of options at one time is cognitively demanding. We design a choice experiment in which respondents are shown a scenario of three hypothetical jobs described in terms of different levels of the same attributes and labeled Job A, Job B and Job C. Respondents are asked which they think is the best job and which they think is the worst job. Each respondent is asked this question for eight different scenarios.

The hypothetical jobs focus on the first job as a registered nurse. The job attributes are based on the 'magnet hospital' literature (Naude and McCabe, 2005; Seago et al., 2001) describing the job characteristics influencing nurses' acceptance of jobs and intention to remain with an employer. The attributes and levels are presented in Table III; the experiment includes 12 attributes, 11 with two levels and one (salary) with four levels. The attributes are appropriate in the context of an entry level job in a new graduate program. In particular, job options are limited to hospitals, as almost all new graduates are employed in hospitals which offer a 'new graduates program'. The 12 attributes cover salary and non-pecuniary aspects including those likely to be relevant to new graduates, including for example clinical rotations, i.e. the opportunity to spend a period of time in different clinical specialties. The attributes were tested in a pilot study with 60 second year nursing students. The pilot study feedback indicated that respondents generally found the scenarios to be understandable and appropriate. In the DCE, attributes were represented by a shortened name and each choice set had a link to an explanatory glossary; see Table III.

We now briefly describe the design underlying the attribute levels. The choice sets are constructed by determining an initial set of 16 jobs which form a resolution 3 fractional factorial design. The other two options in each choice set are then determined by the addition of two generators, chosen so that the resulting set of 16 choice sets of size 3 is D-optimal for the estimation of main effects under the null hypothesis that all of the coefficients are equal to 0. We construct two sets of 16 choice sets using this technique, using two different resolution 3 fractions (so that a larger proportion of the sample space is covered). These sets of 16 choice sets are subdivided into two versions of 8 choice sets and respondents are randomised to one of these 4 versions. A sample choice set of three hypothetical jobs is shown in Figure 1. The full set of 32 choice sets (in coded form), subdivided into the 4 versions of 8 choice sets each, appears in Table I.

3 Sample Description

To become a registered nurse in Australia, students must complete a three-year, university-based degree. Our sample was recruited from the Bachelor of Nursing (BN) degree student enrolment during 2008-2010 at two large Australian universities; one located in a major city, the University of Technology Sydney, and the other located in a regional centre, the University of New England. The sample consists of nursing students in each

year of the course, and new graduates (within 12 months of completing their university course).[2] Student intake includes school-leavers, mature age entry and other nursing workers, seeking to upgrade qualifications. Therefore the sample covers a range of age groups, stages of household formation and exposure to nursing work.

Although the work is part of a broader longitudinal study of nurses' training and job choices, the analysis in this paper is based on the first wave of the survey as these are the only data available to date. The data come from an online survey completed between September 2009 and September 2010, and the analysis focuses on job preferences derived from responses to the DCE component of the survey. The research was conducted in accordance with the Australian Government's National Statement on Ethical Conduct of Human Research and was approved by the research ethics committees at both universities.

Of the 526 respondents, nearly 14% had graduated at the time of survey completion. The majority of respondents were female, born in Australia, aged less than 25 years and reported their health as 'very good' or 'excellent'. Almost one third of the sample lived with a spouse or partner and 16% had dependent children; 49% were still living with their parents all or part of the time. While 65% of the sample had paid work, 35% were employed in health care. Of the 72 graduates, 50 (69%) were employed as a nurse, 11 (15%) were employed in another occupation and 11 (15%) were not in the paid workforce. Among the 454 current students, 63% were employed and 30% were employed as an enrolled nurse or assistant in nursing. More details are provided in Table II.

4 Model specification and selection

In this section of the paper we discuss the various econometric models used to estimate the preference parameters and their performance given our analysis sample. The results are interpreted and discussed for selected models in the following section of the paper. The underlying model is the random utility model (RUM) as developed in Marschak (1960) and McFadden (1981) among others:

$$U_{ij} = x'_{ij}\beta + \epsilon^0_{ij} \tag{1}$$

where U_{ij} denotes the utility associated with an alternative or choice j for person i, (the dependence on the scenario is suppressed) x is a vector of observable characteristics (including an alternative-specific constant), β is a vector of associated utility weights (we discuss heterogeneous coefficients below) and ϵ^0 is a component of utility unobserved by the researcher. The variance of ϵ^0_{ij}, denoted σ^2, is not identified in this model and the estimated parameters β are in fact scaled versions of the true underlying utility weights $\check{\beta}$: $\beta = \check{\beta}/\sigma$. This is the well-known scaling problem as discussed by Louviere and co-authors in various works (for example, see Ohler et al. (2000)).

The most common model of the stochastic process assumes that ϵ^0_{ij} is independent across i and j and is distributed according to an extreme type I (or Gumbel) distribution. This leads to the multinomial logit model (MNL). One of the advantages of the MNL lies in the closed-form representation of the choice probabilities. The probability of individual i choosing alternative k from J possible choices can be written as:

$$\frac{e^{V_{ik}}}{\sum_{j=1}^{J} e^{V_{ij}}} = \frac{e^{x'_{ik}\beta}}{\sum_{j=1}^{J} e^{x'_{ij}\beta}} \tag{2}$$

---

[2]Three respondents were between 12 and 16 months of completing their university degree.

The left most column of results in Table IV presents MNL coefficients for the sample of 12624 observations involving 526 individuals. Four levels were used for the salary in the DCE questionnaire. Specification tests conducted on the MNL and other specifications showed that a linear function of salary did not capture preference weights adequately while a concave function could not be rejected in favor of an unrestricted function of the four salary levels. The ln(salary) is used in all specifications presented below to capture this concave relationship.[3] We note that in this context, alternative-specific constants do not have a natural interpretation since A, B and C are merely labels. These constants (especially that for Job B relative to Job C) are significantly different from zero in some but not all specifications. We discuss this issue further below.

As described above, the DCE component of the survey analysed in this paper asks respondents to choose the best and worst job from a set of three options. This provides a ranking across the three alternatives and allows the use of rank ordered models as well as the usual multinomial specification. The main advantage of a rank ordered model is the gain in efficiency it provides. For each scenario presented to an individual, a full ranking is obtained rather than one preferred choice. Consider 3 possible alternatives: $\{A, B, C\}$. In the context of the RUM, a preference ranking: $A \succ B \succ C$ corresponds to the case where $U_A > U_B > U_C$ and three inequalities characterise the observed ranking instead of the two inequalities that characterise the first best. Note that in the standard rank ordered model, there is only one choice situation and all utilities are known by the individuals before they determine their ranking. In other words, there is no sequential aspect to this ranking.[4]

In a rank ordered logit (ROL), the error term is again assumed to be independently and identically distributed across i and j according to a Gumbel distribution. The probability of observing a ranking, say $A \succ B \succ C$ can be written as:

$$\frac{e^{x'_B \beta}}{e^{x'_B \beta} + e^{x'_C \beta}} \times \frac{e^{x'_A \beta}}{e^{x'_A \beta} + e^{x'_B \beta} + e^{x'_C \beta}} \tag{3}$$

where the person-specific subscript has been omitted. See **?** for a derivation of this equation.

The efficiency gained with rank ordered data depends on the assumption of constant preference parameters over the ranking of alternatives. Some have argued that while the utility weights may remain constant over choices in a single ranking, the variance of the error is likely to increase as one is asked to rank less preferred alternatives. Specifically, assume that the error term attached to the choice of the best alternative among the three possible choices, ϵ^1_{ij}, has a variance equal to σ^2_1 while the error term attached to the choice of the best among the remaining two alternatives has a variance of σ^2_2. As before, errors are assumed to be i.i.d. according to a Gumbel distribution. This leads to the heteroskedastic version of the model developed in Hausman and Ruud (1987) and denoted henceforth as HROL. The probability of $A \succ B \succ C$ can now be written as:

$$\frac{e^{x'_{iB} \beta}}{e^{x'_{iB} \beta} + e^{x'_{iC} \beta}} \times \frac{e^{x'_{iA} \beta \tilde{\sigma}}}{e^{x'_{iA} \beta \tilde{\sigma}} + e^{x'_{iB} \beta \tilde{\sigma}} + e^{x'_{iC} \beta \tilde{\sigma}}} \tag{4}$$

---

[3]A quadratic function performed slightly better than the log transformation but the differences were quantitatively unimportant and other coefficients were not affected. We chose the log function due to the simplification it affords when manipulating and interpreting results. Details are available upon request.

[4]An alternative interpretation is that the individual receives a draw of the unobserved component of his/her utility. In this case the individual's state of mind varies randomly from one choice situation to another but there is still only one draw of the unobserved utility component involved in a single ranking (see McFadden (1981), page 205).

where $\tilde{\sigma} = \sigma_2/\sigma_1$. Note that the sequence of choices now matters and in what follows we assume that individuals choose the best out of three alternatives first followed by the worst option out of the remaining two choices. This is consistent with the presentation of the problem to respondents (see Figure.... discussed in the previous section of the paper).

Table IV presents the results of rank ordered logits with and without heteroskedasticity. A comparison of the estimated coefficients in the MNL and the rank ordered logit without heteroskedasticity shows that preference weights are smaller when including the information on the complete ranking. (The only exceptions are the coefficient on 'Parking' and the constant term on Job A.) This is what we would expect if individuals have higher variance in their choices over less preferred alternatives. It can be seen more clearly when the two components of the ranking are estimated separately. The MNL logit estimates represent the choice of the best job from the three alternatives; the column entitled Logit 2 represents the choice of the best job from the remaining two alternatives after the preferred job is removed[5]. A comparison of these two columns shows that indeed, all coefficients are reduced in size when dealing with the second choice (with the exceptions of the coefficient on 'Parking'and the constant term on Job A.)

The HROL model takes into account the shift in parameters across the two decision nodes in a restricted manner, namely with the scaling parameter $\tilde{\sigma}$. The estimate of $\tilde{\sigma}$ is 1.782 with a standard error of 0.097; in other words, the variance in the second part of the ranking is over three times $(1.782^2 = 3.176)$ that in the choice of the best out of three. The hypothesis of equal variance ($\tilde{\sigma}=1$) is rejected at all conventional levels (the p-value < 0.001) a further indication that the variance of the error increases when ranking less preferred outcomes. A likelihood ratio test (treating the pseudo likelihoods as true likelihoods) rejects the ROL in favor of the unrestricted model where all utility weights are allowed to change (the χ^2 statistic is 168.98 with 14 degress of freedom generating a p-value < 0.001). The choice between the fully unrestricted model and the HROL is not so clear. The AIC would lead to a preference for the unrestricted model (12303.118 vs. 12317.931) while the BIC suggests the opposite (12505.856 vs 12437.244). When examining the parameter estimates, the difference between the MNL and the HROL is small (coefficients differ at the second decimal point only) and no qualitative results are affected. In summary, these results suggest that to incorporate the full ranking, heteroskedastic or possibly more general models should be used.

We now move on to models with heterogeneous utility parameters. Linearity in the deterministic component of the utility and independence across individuals are maintained assumptions. The utility function becomes:

$$U_{ij} = x'_{ij}\beta_i + \epsilon^3_{ij} = x'_{ij}(\tilde{\beta} + \eta_i) + \epsilon^3_{ij} \tag{5}$$

It is assumed that individuals know their utility weights β_i's and draws ϵ^3_{ij}'s but these are not observed by the researcher. The mixed logit is derived from this model under the assumption that the ϵ^3_{ij}'s are independently drawn from the Gumbel distribution. The unconditional probability of observing a choice say k can be written as:

$$\int \left(\frac{e^{x'_{ik}\beta_i}}{\sum_{j=1}^{J} e^{x'_{ij}\beta_i}} \right) f(\beta_i) \; d\beta \tag{6}$$

where $f(.)$ is the joint density of the vector β_i. Following most of the literature we assume

---

[5]With only two alternatives, the estimates corresponding to the best job are simply the negative of the estimates for the choice of the worst job.

that the mixing distribution $f(.)$ is normal: $\beta_i \sim \text{MVN}(\tilde{\beta}, \Sigma)$.[6] The resulting probabilities do not have a closed form and must be simulated in the estimation. The method of maximum simulated likelihood is used.[7] There are several choices to be made in this model; particular elements of the parameter vector β_i can be fixed across individuals (i.e. $\beta_i = \tilde{\beta}$ for some of the attributes) and covariances across random parameters can be set to zero to simplify the estimation. Since we are interested in patterns of tastes, it is natural to adopt a general specification initially. However, an unrestricted joint normal mixing distribution could not be estimated adequately. Convergence was often reached but the correlation parameters did not stabilise even after a large number of replications.[8] In what follows we present estimates where correlations in utility weights across attributes are set at zero.[9] In the mixed logit model, parameters stabilised after 10,000 replications.[10]

The left-most columns of Table V present estimates for the means and the standard deviations of the vector β_i based on 10,000 replications. The means of the distribution of attribute weights are all significantly different from zero at a 1% level of significance except for 'abundant parking'. This follows patterns in the multinomial logit with fixed utility weights. Both alternative-specific constants are significantly different from zero; this raises the possibility that certain choices are made based on criteria other than the attributes of the jobs. All standard deviations for the attribute weights are significantly different from zero, an indication of heterogeneity in the utility weights across individuals. A comparison of the AIC and BIC measures also supports the use of the mixed model over the MNL.

The use of the more general mixed logit yields estimates for the means $\tilde{\beta}$'s that are almost twice as high as the corresponding MNL results (ignoring the alternative specific constants). As described by Revelt and Train (1998) a scaling up of coefficients is to be expected as the unexplained component is likely to have a smaller variance in the MXL (hence cause a smaller scaling down of the attribute weights) since it excludes variation due to preference heterogeneity in the weights. In terms of relative importance however there is not much difference in the utility weights. Ignoring the alternative-specific constants, there are two changes in the ranking of attributes: 'appropriate responsibility' has moved ahead of 'flexible rostering' and 'nurses encouraged'; and 'well equipped' and 'well staffed' have switched ranks. These estimates are fairly close together so a switch in their ranking is not that surprising and overall qualitative results are similar in the two models. (This will be more apparent in the next section of the paper.)

The right-most columns of Table V provide results for the generalized multinomial logit (GMNL) model recently developed in Fiebig et al. (2010) and Keane and Wasi (2009). In the GMNL, the distributional assumptions (along with the panel nature of

---

[6]There is some debate over the use of a normal density for the parameters attached to a monetary value such as the salary variable. Some researchers force the weight to be positive for all individuals by specifying a log normal density for such parameters. There is disagreement in the literature regarding the impacts of such assumptions (see Greene and Hensher (2003)). We use the normal for all parameters.

[7]In what follows Stata version 11 is used to estimate the simulated likelihood. Halton draws are used and 43 initial draws are burned (see Train (2009)).

[8]The use of the rank ordered data did not help in identifying the correlations in the fully unrestricted model.

[9]See Train (2009), pp. 140-141 for a discussion of the difficulty in identifying correlations in models with many attributes.

[10]There is no conscensus in the literature on an acceptable level of variation in parameters across simulated likelihood estimates. We chose 10% as a maximum amount of variation allowed in any of the means or standard deviations in the mixture distribution. An alternative strategy is to use one standard deviation as a maximum amount of variation allowed in parameters; Walker (personal communication). Our choice of 10,000 replications satisfies both of these criteria.

the data) are used to identify parameters of the distribution of the scaling factor as well as how it interacts with the utility weights. This model can be seen as a generalization of the MXL in which the variance of the error term is heterogeneous across individuals. Specifically, the utility function in the GMNL is written as

$$U_{ji} = x'_{ji}(\varsigma_i \tilde{\beta} + \eta_i^*) + \epsilon_{ji}^4 \qquad (7)$$
$$= x'_{ji}(\varsigma_i \tilde{\beta} + \gamma \eta_i + (1 - \gamma)\varsigma_i \eta_i) + \epsilon_{ji}^4 \qquad (8)$$

where ς_i is an individual-specific scalar, unobserved by the researcher and known by the individual, scaling the $\tilde{\beta}$ vector up and down, γ is a parameter that allows η_i to be scaled up by ς_i (when $\gamma = 0$) or to vary independently (when $\gamma = 1$). In the most common version of GMNL, ς_i is assumed to follow the lognormal distribution, $\ln(\varsigma_i) \sim N(\bar{\varsigma}, \tau^2)$ with $\bar{\varsigma}$ normalized to $-\tau^2/2$. Initially, γ was restricted to the (0,1) interval but this restriction was abandoned following a discussion in Keane and Wasi (2009).

Initial estimates of the GMNL model with unrestricted γ yielded an estimate of γ equal to 0.099 with a standard error of 0.175; hence there is no support for differential scaling of the mean and the heterogeneous component of the attribute weight.[11] The left-most columns of Table V present estimates for the GMNL model with γ fixed at 0. The estimated standard deviation of the log of the scaling factor τ is highly significant; evidence of heterogeneity in the scale is found in these data. The simulated likelihood is improved in GMNL relative to the MXL as are both AIC and BIC statistics. In terms of the qualitative results, the GMNL yields means and standard deviations that are higher than their mixed logit counterparts but the ranking across attributes is unchanged.

The next set of results correspond to a heteroskedastic rank ordered version of the GMNL model (HROGMNL) that uses information on the ranking across the 3 jobs. As above, we assume that the individual chooses the best out of three alternatives first and the best (or worst) out of the remaining 2 alternatives second; also we allow for a shift in the mean of the scaling factor between these two decision nodes. This corresponds to an extension of the HROL model described previously to a framework with heterogeneity in the utility weights and in the scaling factor. Specifically, ς_i is assumed to follow the lognormal distribution, $\ln(\varsigma_i) \sim N(\bar{\varsigma}, \tau^2)$ with $\bar{\varsigma} = -\tau^2/2 + \delta * S$ with S equal to 0 for the first decision node and 1 for the second choice in the ranking. In other words, δ measures the shift in the mean of $\ln \varsigma$ as respondents move from their first best to their second best choice, a shift which is assumed to be common to all individuals.[12]

The estimates for the heteroskedastic rank ordered GMNL model are presented in the left-most columns of Table VI. Based on previous results γ is fixed at zero. It is interesting that after allowing for individual heterogeneity in means and scaling, there is no evidence of a shift in the scaling factor across choice nodes; i.e. δ is small and insignificant. For most attributes the mean of the attribute weight is smaller in this specification but the ranking is similar to that obtained previously. We also note that although there is evidence of heterogeneity in the job specific constants (the standard deviations are significantly different from zero at a 1% level of significance) their means are small and insignificant at 1% in this model. Forcing the job specific constants to be the same (for the same individuals) in the two decision nodes gets rid of most of the preference of Jobs A and B over C. Most importantly, the means and standard deviations of the attribute weights are very similar in this model compared to the previous estimations; even if certain individuals use a rule of thumb (such as the middle position of the job on

---

[11] Detailed results are available from the authors.
[12] This is the first estimation of such models that we are aware of.

9

the screen) to help in their choices, the results involving the relative importance of job attributes are not greatly affected.

As a final robustness check, we use information on the response time as a proxy for motivation or interest of the respondent and focus on a subset of the sample who completed the survey within a relatively short time-frame. Specifically, we construct a variable equal to the difference between the date at which the link for the survey website was sent to the respondent and the date of the survey completion. The variable is referred to as "response time" and is measured in days. The mean and median number of days elapsed between the sending of the link and the completion of the survey are 31 and 5 days respectively; the range is 0 to 340 days. Although the majority of the 526 individuals answered within one week of receiving the link to the survey website, a substantial number also took a long time: 225 individuals (43%) waited over 2 weeks and of these 135 (26% of the total sample) waited more than 50 days before completing the survey. The right most columns of Table VI provide results for a mixed logit estimated on the restricted sample which excludes all individuals who answered more than 10 days after receiving the link to the survey website. The remaining sample numbers 6768 observations or 54% of the original sample and involves 282 individuals. As shown in Table VI, the mean attribute weights for the reduced sample are higher compared to the MXL results on the total sample with one exception ("appropriate responsibility"). This could be explained by the reduced sample having a smaller variance in the unexplained component of the utility. However, the differences between the estimates are very small and except for salary, are found at the second decimal point only (for salary the difference is still less than 10% of the estimate).

5   Interpretation of estimation results

Two sets of figures are computed from the estimation results to make the figures easier to interpret: predicted probabilities of job choice and willingness to pay measures. The predicted probabilities answer the following question: 'What is the change in the predicted probability of choosing a job Z instead of another job Y if the only difference in the two jobs lies in the level of attribute k?' For the multinomial logit this can be written as:

$$Prob\{U_Z > U_Y\} = Prob\{x'_Z \beta - x'_Y \beta > \epsilon^0_Y - \epsilon^0_Z\} = \left(\frac{e^{\beta_k}}{1 + e^{\beta_k}}\right) \qquad (9)$$

where it is assumed that the jobs differ only in the attribute k and that this attribute shifts by one unit (we discuss the shift in salary below). The base job Y is defined as the worst possibility in the sense that all attributes are set at their least preferred levels. The resulting predicted probabilities will be > 0.5 since β's > 0. (The predicted probability will equal 0.5 if the attribute is unimportant ($\beta_k = 0$) and the choice is hence completely random.)[13]

Table VII presents predicted probabilities for the main models in our analysis. The figures in the table measure the predicted probability of accepting a job in which the corresponding attribute has shifted to its preferred level all other job attributes held fixed at their base level. For the salary the shift is from 800 to 1250 dollars per week; all other attributes are binary and the shift is from zero to one. All predicted probabilities

---

[13]When the coefficients are random and normally distributed, the predicted probability has a logit-normal distribution. The mean of this distribution has no analytical solution in general but the median is well-defined and equal to the logistic function evaluated at the mean $\tilde{\beta}_k$. Hence, although technically the statistic is different, qualitative interpretations are similar.

are significantly different from 0.5 at the 1% level except for those corresponding to "Parking".

Salary has the highest effect on the predicted probability; when salary shifts from 800 to 1250, an individual is almost sure to choose the new job over the old one (the probability is over 90% in all models except for the MNL model where the probability is 77%). Only an extreme value for the unobserved components of utility would lead to a preference for the original job. We can form roughly four groups of attributes based on their importance: salary, supportive management/staff and quality of care; appropriate responsibility, flexible rostering, encouragement; well equipped and well staffed premises; public hospital, 3 rotations, flexible hours and abundant parking. The ranking across these groups is robust across all models; indeed the ranking within the groups is also the same across models with only a few exceptions.

An alternative approach transforms utility weights into dollar values; specifically, willingness to pay (WTP) measures are constructed as marginal rates of substitution (MRS) between an attribute and a monetary attribute, in our case salary. This statistic answers the following question: 'What is the loss in salary that would keep utility constant when one attribute, say k, is shifted to its preferred level, all other attributes remaining unchanged?'[14] Denote the coefficient on ln(salary) as β_s and change attribute k from 0 to 1:

$$\Delta U = 0 \Rightarrow \beta_k + \beta_s \ln(m \times \text{salary}) = \beta_s \ln(\text{salary}) \tag{10}$$

where m is the proportion of the salary which is retained and which guarantees constant utility. Measuring the loss in salary in dollars from the base of 800 yields

$$\text{WTP} = 800 \times (1 - m) = 800 \times (1 - e^{-\beta_k/\beta_s}). \tag{11}$$

When coefficients are fixed, it is straightforward to derive estimates for WTP by using point estimates for β. In the mixed logit and its extensions, the attribute weights are normally distributed variables and their ratio will have a Cauchy ratio distribution. For general parameter values, the mean of this ratio is not well-defined. The median exists for all values of the parameters of the distributions and it has a well-defined pdf; however, in general the pdf does not have a closed form representation and must be simulated.

Table VIII presents marginal rates of substitution (in absolute value) between the attributes and salary. WTP figures are in dollar values and should be compared to a base salary of $800 per week. In the MNL, point estimates are used to evaluate the MRS. In models with individual heterogeneity in the attribute weights, two estimates are provided. In the top panel, WTP is measured with the coefficients set at their mean values. Although the ratio does not exist for all values of the denominator, it is still a useful estimator of the WTP. When comparing with MNL, differences in this estimate of the WTP will be due to differences in the estimated mean attribute weights only. As we can see from the table, the ranking of the attributes and indeed the dollar values placed on the attributes are very similar across models; even though the average attribute weights are shifted up in the random coefficient models, they are shifted up in a systematic way and the WTP measures are only minimally affected. Standard errors are computed with the delta method (not shown). For all models, only parking has a willingness to pay which is not significantly greater than zero at the 1% level.

In panel b, the distributions of the random coefficients are taken into account when computing the WTP measures; specifically, the distributions of the WTP measures are

---

[14]We are using willingness-to-pay in a restricted sense; this experiment does not yield welfare measures that can be applied in arbitrary situations since they do not allow for a nurse's choice to move out of nursing jobs altogether (see Lancsar and Savage (2004) for more details).

simulated with 100,000 replications. The median of the simulated distribution along with the 25th and 75th percentiles are provided. The parameter values are such that the median WTP is scaled down substantially relative to the WTP at mean coefficients; however the rankings are the same with but a few shifts within the 4 groups of attributes. Figure 2 presents the 25th percentile, the median and the 75th percentile for the simulated WTP distributions. The underlying means and standard deviations for the attribute weights are estimated with the GMNL model.

The simulated WTP distributions show a large amount of dispersion in the weights placed on job characteristics. This reflects the estimated standard deviations around mean attribute weights. For the first seven attributes (salary, to well staffed), the ratio of the mean to the interquartile range is generally ≥ 0.5 while the figure for the remaining attributes is normally ≤ 0.25. Interestingly, the first group of attributes have clear better and worse levels; for example a higher salary is always better, excellent care is better than low quality of care, and so on. Our respondents may have different strengths of preferences, but a well equipped hospital is generally preferred to a poorly equipped one. In contrast, the characteristics in the second group do not have clear better or worse levels. With these attributes individuals have quite divergent preferences, with some seeing them as positive contribution to utility while others consider the same attribute as having a negative impact. For example, 3 rotations will be positive for those nurses who wish to experience a variety of clinical areas; but equally it will have a negative impact for those nurses who are already certain they want to work in one field of nursing. This preference diversity, as opposed to strength, is an important issue to be considered in designing policies to improve retention.

6   Job preferences and time in the program

In this section of the paper we investigate if relative weights placed on job attributes differ with the progression through the program of study and the initial post-graduation experience with the workplace. Without panel data we cannot control for unobserved individual characteristics that may differ across the subsamples by year of program; nevertheless since our cross-section data spans the whole length of the program of study we can investigate the possibility of systematic differences in attribute weights for individuals at different levels in the program. Specifically, we construct dummy variables to represent the respondent's year in the program. In total there are 4 groups: 1st year, 2nd year, 3rd year (including any 4th year) and graduates. The distribution of the 526 individuals is as follows: 183 (35%) in first year, 137 (26%) in second year, 134 (25%) in third year and 72 (14%) graduates. We estimate a MXL where all attribute weights are heterogeneous across agents and where the means of the distributions shift across years in the program.

In Table IX estimates are presented for a MXL where all attribute weights are heterogeneous across agents and where the mean of the distribution shifts across years in the program. Hence, the distributions of the attribute weights are shifted horizontally across the years in the program. (Say something about gmnl estimates.) The first column presents the means of the attribute weights for the first year students. The next three columns present differences from the year 1 mean weight. The right-most column presents the standard deviations of the distributions of the attribute weights (these are assumed constant across years in the program.)[15]

---

[15]The job specific constants are assumed to have a fixed distribution across the years.

These results show that although there is a large amount of stability in relative weights over the years in the program there is also some shifting in relative importance of job characteristics. Joint tests show that year one mean attribute weights are jointly significantly different from those of the other students. P-values of joint significance tests on the differences in mean attributes across years are 0.014 for a test between year 1 and graduates; 0.041 between years 1 and 3; and 0.079 between years 1 and 2.[16] Tests on individual attributes show that equality of mean attribute weights across the 4 groups of respondents is rejected for three attributes (based on a 10% level of significance): 3 rotations (p-value of 0.010), flexible rostering (p-value of 0.008) and quality of care (p-value of 0.0327). In addition, several shifts in mean attributes are individually significantly different from zero.

Predicted probabilities of job choice and willingness-to-pay measures are provided in Table X.[17] We present figures for year 1 and shifts in the figures for subsequent years. We also present the ranking of the mean attribute weights for year 1 students and graduates to show that shifts occur in the relative ranking of the attributes as well as in the magnitude of the attribute weights.

Briefly, graduates place more weight on 3 rotations and flexible hours and less weight on quality of care relative to first years. The third year group also places more weight on appropriate responsibility and flexible rostering relative to first year. What differs as nurses move through their education and into the nursing workforce? Trainees in their later years and then graduate nurses have gained more clinical experience and insights, and are older than their first year counterparts. Our findings suggest that this greater clinical understanding results in greater weight placed on appropriate responsibility, and that the realities of working shift work and/or changing family situations explain the stronger preference for flexible hours.[18]

7   Conclusions

This paper is the first study of nurses job preferences that applies DCE methods to a developed country workforce. It adds to the previous literature on stated intentions to quit, as those studies are limited to comparing the job characteristics of actual jobs with unknown alternatives. In contrast, DCEs allow the construction of a much wider range of hypothetical alternatives with defined attributes, and thus let us explore more fully how different policy options would impact attrition and retention. Our DCEs use a greater range of job attributes than previous studies, thus increasing the realism of the choice scenarios. The choice of attributes reflects factors that have been shown to be important to nurses in various literatures and the levels of the attributes have been chosen to make the jobs realistic in the context of our sample.

This paper is also the first to focus on the transition through university training and into the labour force. Our sample comprises students at different stages of training and new graduate nurses; this is a particularly interesting group since junior nurses on average have the lowest retention levels in the profession. We find that while preferences are similar over the transition, for nurses in their first job, supportive management/staff is valued significantly more than for student nurses. Indeed, in terms of ranked order, it is more important than salary (at normal levels). Having appropriate levels of responsibility

---

[16]Using a 10% level of significance, the only other pair-wise comparison to yield jointly significant differences in mean attributes is that involving year 2 and graduates where the p-value is 0.064.

[17]To simplify, WTP measures are computed at the mean of the attribute weight.

[18]In a companion paper, we explore shifts in preferences across observable personal chatacteristics.

and a greater range of training (number of rotations) are also ranked more highly by new nurses than students.

The paper makes a methodological contribution in that we adapt state-of-the-art models of heterogeneity (MXL and GMNL) to best-worst information and allow for heteroskedasticity across choice nodes. Thus we allow for flexible unobserved heterogeneity in preferences and possible shifts in scale across the best-worst choices. Our results remain remarkably robust across models (even in very flexible frameworks) and suggest that although there is significant scale heterogeneity, there is no evidence of systematic shifts in scale across best-worst choices.

The policy implications of our results are several. First, salary remains an important factor in making nursing jobs attractive. Although non-pecuniary benefits are also important, policy should not ignore pay levels for nurses. Along with salaries, policies which promote a supportive workplace culture and high quality of care will also be effective in making nursing jobs more attractive. Second, there is evidence of substantial heterogeneity of preferences; attributes that make jobs more attractive for some nurses can be disliked by others. Nursing retention could be improved by designing quite different employment packages to appeal to these different tastes. This represents a shift in policy, particularly in those countries such as Australia with a centralised approach to setting salaries and employment benefits. Third, we see that the transition from university student to new graduate nurse is apparently a time when a supportive workplace culture and the level of responsibility make a difference, so that policies which lessen the stress and possible feelings of isolation may also be important in retaining the vulnerable group of new graduates. Our study is designed as a panel and future work will report on how different nursing experiences affect preferences and retention.

References

Aiken, L., Clarke, S., Sloane, D., Sochalski, J., and Silber, J. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*, 288:1987–93.

Antonazzo, E., Scott, A., Skatun, D., and Elliott, R. F. (2003). The labour market for nursing: a review of the labour supply literature. *Health Economics*, 12(6):465–78.

Barron, D. and West, E. (2005). Leaving nursing: an event-history analysis of nurses' careers. *J Health Serv Res Policy*, 10:150–157.

Blaauw, D., Erasmus, E., Pagaiya, N., Tangcharoensathein, V., Mullei, K., Mudhune, S., Goodman, C., English, M., and Lagarde, M. (2010). Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment. *Bulletin of the World Health Organization*, 88(5):350–356.

Casey, K., Fink, R., Krugman, M., and Propst, J. (2004). The graduate nurse experience. *Journal of Nursing Administration*, 34(6):303–311.

Cunich, M. and Whelan, S. (2010). Nurse education and the retention of registered nurses in new south wales. *Economic Record*, 85(274):396–413.

Di Tommaso, M. L., Strom, S., and Saether, E. M. (2009). Nurses wanted: Is the job too harsh or is the wage too low? *Journal of Health Economics*, 28(3):748–757.

Doiron, D, ., Hall, J., and Jones, G. (2008). Is there a crisis in nursing retention in new south wales? *Australia and New Zealand Health Policy*, 5:19.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science*, 29:393–421.

Flynn, T. N., Louviere, J. J., Peters, T. J., and Coast, J. (2007). Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*, 26(1):171–189.

Fochsen, G., Josephson, M., Hagberg, M., Toomingas, A., and Lagerstrom, M. (2006). Predictors of leaving nursing care: a longitudinal study among swedish nursing personnel. *Occupational and Environmental Medicine*, 63:198–201.

Frijters, P., Shields, M. A., and Price, S. W. (2007). Investigating the quitting decision of nurses: panel data evidence from the british national health service. *Health Economics*, 16(1):57–73.

Greene, W. H. and Hensher, D. A. (2003). The mixed logit model:the state of practice. *Transportation*, 30:133–176.

Greene, W. and Hensher, D. (2010). Does scale heterogeneity across individuals matter? an empirical assessment of alternative logit models. *Transportation*, 37(3):413–428.

Hall, J., Fiebig, D., King, M., Hossain, I., and Louviere, J. (2006). What influences participation in genetic carrier testing?: Results from a discrete choice experiment. *Journal of Health Economics*, 25(3):520–537.

Hausman, J. and Ruud, P. (1987). Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, 34:83–104.

Heinz, D. (2004). Hospital nurse staffing and patient outcomes: a review of current literature. *Dimensions of critical care nursing*, 23:44–50.

Keane, M. and Wasi, N. (2009). Comparing alternative models of heterogeneity in consumer choice behavior. Technical report, University of Technology, Sydney.

King, M. T., Hall, J., Lancsar, E., Fiebig, D., Hossain, I., Louviere, J., Reddel, H. K., and Jenkins, C. R. (2007). Patient preferences for managing asthma: results from a discrete choice experiment. *Health Economics*, 16(7):703–717.

Lagarde, M. and Blauuw, D. (2009). A review of the application and contribution of discrete choice experiments to inform human resources policy interventions *Human Resources for Health*, 7:62 doi:10.1186/1478-4491-7-62.

Lancsar, E. and Savage, E. (2004). Deriving welfare measures from discrete choice experiments: Inconsistency between current methods and random utility and welfare theory. *Health Economics Letters*, 13:901–907.

Louviere, J., Finn, A. and Timmermans, H. (1994). Retail Research Methods. *Handbook of Marketing Research*, 2nd Edition, McGraw-Hill, New York.

Mangham, L. J. and Hanson, K. (2008). Employment preferences of public sector nurses in malawi: results from a discrete choice experiment. *Tropical Medicine and International Health*, 13(12):1433–41.

Marschak, J. (1960). Binary choice constraints on random utility indications. In Arrow, K., editor, *Stanford Symposium on Mathematical Methods in the Social Sciences*, pages 312–329. Stanford University Press, Stanford, CA.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. The MIT press, MIT.

Naude, M. and McCabe, R. (2005). Magnet hospital research pilot project conducted in hospitals in western australia. *Contemporary Nurse*, 20:38–55.

Needleman, J. and Hassmiller, S. (2009). The role of nurses in improving hospital quality and efficiency: Real-world results. *Health Affairs*, 28:625–633.

Nooney, J. G., Unruh, L., and Yore, M. M. (2010). Should i stay or should i go? career change and labor force separation among registered nurses in the u.s. *Social Science and Medicine*, 70(12):1874–1881.

Ohler, T., Le, A., Louviere, J., and Swait, J. (2000). Attribute range effects in binary response tasks. *Marketing Letters*, 11:249–260.

Oulton, J. (2006). The global nursing shortage: An overview of issues and actions. policy, politics. *Nursing Practice*, 7:34S–39S.

Penn-Kekana, L., Blaauw, D., San Tint, K., Monareng, D., and Chege, J. (2005). Nursing staff dynamics and implications for maternal health provision in public health facilities in the context of hiv/aids. Technical report, Frontiers in Reproductive Health, Population Council.

Productivity Commission (2005). *Australia's health workforce*. Research Report, Canberra.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *The Review of Economics and Statistics*, 80:647–657.

Rother, J. and Lavizzo-Mourey, R. (2009). Addressing the nursing workforce: A critical element for health reform. *Health Affairs*, 28:620–624.

Seago, J., Ash, M., Spetz, J., Coffman, J., and Grumbach, K. (2001). Hospital registered nurse shortages: Environmental, patient, and institutional predictors. *Health Services Research*, 36:831–52.

Shields, M. A. and Ward, M. (2001). Improving nurse retention in the national health service in england: The impact of job satisfaction on intentions to quit. *Journal of Health Economics*, 20(5):677–701.

Shields, M. A. (2004). Addressing nurse shortages: What can policy makers learn from the econometric evidence on nurse labour supply? *Economic Journal*, 114(499):F464–F498.

Train, K. E. (2009). *Discrete Choice Methods with Simulation, Second Edition*. Cambridge University Press, Cambridge.

Vermeulen, B., Goos, P., and Vandebroek, M. (2010). Obtaining more information from conjoint experiments by best-worst choices. *Computational Statistics and Data Analysis*, 54(6):1426–1433.

Zeytinoglu, I. U., Denton, M., Davies, S., Baumann, A., Blythe, J., and Boos, L. (2006). Retaining nurses in their employing hospitals and in the profession: Effects of job preference, unpaid overtime, importance of earnings and stress. *Health Policy*, 79(1):57–72.

Zeytinoglu, I. U., Denton, M., and Plenderleith, J. M. (2011). Flexible employment and nurses' intention to leave the profession: The role of support at work. *Health Policy*, 99(2):149–157.

Figure 1: Sample choice set with three hypothetical jobs

There are jobs available in three programs for new graduates which have the following characteristics:

To review the features of jobs, please click here.

Choice 7 of 8			
Features of Job	**Job A**	**Job B**	**Job C**
Location	Private hospital	Private hospital	Public hospital
Clinical rotations	None	None	Three
Work hours	Fulltime only	Part-time or fulltime	Fulltime only
Rostering	Inflexible, does not allow requests	Flexible, usually accommodating requests	Inflexible, does not allow requests
Staffing levels	Usually well-staffed	Frequently short of staff	Usually well-staffed
Workplace culture	Unsupportive management and staff	Unsupportive management and staff	Supportive management and staff
Physical environment	Poorly equipped and maintained facility	Poorly equipped and maintained facility	Well equipped and maintained facili
Professional development and progression	No encouragement for nurses	Nurses encouraged	No encouragement for nurses
Parking	Abundant and safe	Limited	Abundant and safe
Responsibility	Appropriate responsibility	Appropriate responsibility	Too much responsibility
Quality of care	Poor	Excellent	Excellent
Weekly salary	$1,250	$800	$1,100
Considering these three jobs:			
Which would you MOST like to get?	⚪ Job A	⚪ Job B	⚪ Job C
Which would you LEAST like to get?	⚪ Job A	⚪ Job B	⚪ Job C

>>

Table I: Versions for the Discrete Choice Experiment (in coded levels)

	Version 1			Version 2	
000000000000	001110011011	110001100113	101000101013	100110110000	011001001102
000011011011	001101000002	110010111100	101011110002	100101101013	011010010111
000101101102	001011110113	110100001011	101101000111	100011011102	011100100000
000110110113	001000101100	110111010002	101110011100	100000000111	011111111013
011000011112	010110000103	101001111001	110000110101	111110101112	000001010010
011011000103	010101011110	101010100012	110011101110	111101110101	000010001003
011101110010	010011101001	101100010103	110101011003	111011000010	000100111112
011110101001	010000110012	101111001110	110110000012	111000011003	000111100101

	Version 3			Version 4	
111111111113	110001100100	001110011002	010111010102	011001001113	100110110011
111100100100	110010111111	001101000013	010100001111	011010010102	100101101000
111010010011	110100001002	001011110100	010010111000	011100100011	100011011113
111001001002	110111010013	001000101111	010001100013	011111111000	100000000102
100111100001	101001111012	010110000110	001111001010	000001010001	111110101103
100100111012	101010100003	010101011101	001100010003	000010001010	111101110112
100010001103	101100010110	010011101012	001010100112	000100111103	111011000001
100001010110	101111001101	010000110003	001001111101	000111100112	111000011010

Table II: Sample characteristics of 526 respondents

Characteristic	%
Bachelor of nursing	
Graduate	13.7
1st year student	34.8
2nd year student	26.0
3rd year student	25.5
Age in years	
19 or less	22.2
20-24	39.2
25-29	13.9
30 or more	24.7
Female	89.4
Born in Australia	67.9
Speak English at home	82.9
Household	
Live with parents	49.2
Live with partner/spouse	31.8
Children aged less than 16 years	15.8
Self-rated health	
Very good/excellent	69.2
Good	26.4
Fair/poor	4.4
Gross income*	
Less than $20,000 pa	45.6
$20,000-$39,999 pa	15.2
$40,000-$79,999 pa	13.1
$80,000 pa or more	12.6
Missing	13.5
Government student support**	35.6
Employed	65.2
Employed in nursing	35.0

Notes:

* Income = own and partner's income in Australian dollars 2009-10;

** During final year of study for graduates

Table III: Attributes and Levels for the Discrete Choice Experiment and associated model variable names

Glossary definition of attribute	Attribute name	Levels	Variable
The type of hospital where the new graduate program is located	Location	Private hospital Public hospital	Public hosp
The number of rotations to different clinical areas	Clinical rotations	None Three	3 rotations
Whether the new graduate program offers fulltime and part-time positions, or fulltime only	Work hours	Fulltime only Part-time or fulltime	Flex hours
The flexibility of the rostering system in accommodating requests	Rostering	Inflexible, does not allow requests Flexible, usually accommodating requests	Flex rost
The hospital's reputation regarding staffing levels	Staffing levels	Frequently short of staff Usually well-staffed	Well staff
The hospital's reputation regarding the workplace culture in terms of support from management and staff	Workplace culture	Unsupportive management and staff Supportive management and staff	Supp mgt
The hospital's reputation regarding the physical work environment in terms of equipment and appearance	Physical environment	Poorly equipped and maintained facility Well equipped and maintained facility	Well equip
The hospital's reputation regarding whether nurses are encouraged and supported in professional development and career progression	Professional development and progression	No encouragement for nurses Nurses encouraged	Encourage
The parking facilities	Parking	Limited Abundant and safe	Parking
The hospital's reputation regarding the responsibility given to nurses, relative to their qualifications and experience	Responsibility	Too much responsibility Appropriate responsibility	App resp
The hospital's reputation regarding the quality of patient care	Quality of care	Poor Excellent	Excell care
The gross weekly salary	Salary*	$800 $950 $1,100 $1,250	Salary

*Modelled as a continuous variable.

Table IV: Multinomial and rank ordered models. Standard errors in parentheses.

| | **Models** | | | |
	MNL	ROL	Logit2	HROL
Salary	1.550***	1.235***	0.814***	1.526***
	(0.095)	(0.071)	(0.152)	(0.066)
Supp mgt	1.044***	0.828***	0.513***	1.033***
	(0.049)	(0.039)	(0.047)	(0.040)
Excell care	0.832***	0.731***	0.598***	0.858***
	(0.050)	(0.038)	(0.045)	(0.036)
App resp	0.475***	0.378***	0.233***	0.471***
	(0.048)	(0.035)	(0.043)	(0.036)
Flex rost	0.542***	0.389***	0.206***	0.516***
	(0.042)	(0.031)	(0.042)	(0.036)
Encourage	0.519***	0.438***	0.377***	0.547***
	(0.045)	(0.034)	(0.043)	(0.036)
Well equip	0.374***	0.350***	0.239***	0.392***
	(0.039)	(0.029)	(0.041)	(0.036)
Well staff	0.400***	0.362***	0.272***	0.423***
	(0.037)	(0.029)	(0.044)	(0.034)
Public hosp	0.241***	0.130***	0.048	0.208***
	(0.040)	(0.028)	(0.047)	(0.036)
3 rotations	0.205***	0.148***	0.023	0.174***
	(0.040)	(0.027)	(0.037)	(0.036)
Flex hours	0.128***	0.076***	0.034	0.115***
	(0.035)	(0.027)	(0.039)	(0.034)
Parking	0.064*	0.073***	0.103**	0.087***
	(0.038)	(0.028)	(0.042)	(0.033)
Job B Cst	0.131***	0.084***	0.111**	0.127***
	(0.044)	(0.032)	(0.054)	(0.041)
Job A Cst	0.008	0.014	0.193***	0.052
	(0.046)	(0.030)	(0.051)	(0.041)
$\tilde{\sigma}$				1.782†††
				(0.099)
Sample Size	12624	21040	8416	21040
PLLikelihood	-3492.546	-6208.049	-2631.014	-6143.966
AIC	7013.091	12444.098	5290.027	12317.931
BIC	7117.298	12555.457	5388.558	12437.244

Notes: MNL refers to a multinomial logit, ROL to a rank ordered logit, HROL to a heteroskedastic rank ordered logit and Logit2 is a logit using data on the second choice in the ranking of the three alternatives. The coefficient on salary measures the change in utility caused by moving from a job with a weekly salary of 800 to a job with a weekly salary of 1250. The standard errors are robust to arbitrary heteroskedasticity and to correlations across observations from the same individuals. PLLikelihood indicates a pseudo log likelihood, AIC refers to the Akaike information criterion and BIC to the Bayesian information criterion. *** indicates that the parameter is significantly different from zero at a 1% level of confidence, ** at 5% and * at 10%.†††indicates that the parameter is significantly different from 1 at a 1% level .

Table V: Mixed logit and generalized mixed logit models. Standard errors in parentheses.

	Models			
	MXL		GMNL	
	Mean	St.Dev.	Mean	St.Dev.
Salary	2.883***	2.828***	4.281***	4.073***
	(0.241)	(0.287)	(0.819)	(0.783)
Supp mgt	1.946***	1.381***	2.869***	1.808***
	(0.151)	(0.145)	(0.528)	(0.377)
Excell care	1.438***	1.321***	2.100***	1.741***
	(0.120)	(0.119)	(0.403)	(0.366)
App resp	0.961***	1.024***	1.363***	1.242***
	(0.105)	(0.137)	(0.265)	(0.317)
Flex rost	0.912***	0.851***	1.359***	1.140***
	(0.090)	(0.125)	(0.274)	(0.280)
Encourage	0.822***	0.611***	1.255***	0.846***
	(0.083)	(0.146)	(0.253)	(0.262)
Well equip	0.713***	0.622***	1.055***	0.814***
	(0.084)	(0.157)	(0.215)	(0.251)
Well staff	0.683***	0.549***	1.052***	0.745***
	(0.075)	(0.127)	(0.215)	(0.263)
Public hosp	0.441***	0.748***	0.618***	0.925***
	(0.076)	(0.147)	(0.146)	(0.234)
3 rotations	0.375***	0.795***	0.544***	1.071***
	(0.077)	(0.132)	(0.144)	(0.247)
Flex hours	0.210***	0.578***	0.322***	0.767***
	(0.062)	(0.141)	(0.107)	(0.249)
Parking	0.101	0.421**	0.159	0.652***
	(0.061)	(0.179)	(0.098)	(0.251)
Job B Cst	0.369***	0.117	0.364***	0.216
	(0.095)	(0.467)	(0.099)	(0.183)
Job A Cst	0.244**	0.300	0.242**	0.352**
	(0.096)	(0.205)	(0.100)	(0.170)
τ			0.690***	
			(0.151)	
Sample Size	12624		12624	
SLLikelihood	-3287.217		-3278.283	
AIC	6630.433		6614.566	
BIC	6838.847		6830.423	

Notes: MXL refers to a mixed logit and GMNL to a generalised mixed logit model. For the simulations, 10,000 Halton draws are made after burning the initial 43 draws. The coefficient on salary measures the change in utility caused by moving from a job with a weekly salary of 800 to a job with a weekly salary of 1250. For the MXL the standard errors are robust to arbitrary heteroskedasticity and to correlations across observations from the same individuals. SLLikelihood indicates a simulated log likelihood, AIC refers to the Akaike information criterion and BIC to the Bayesian information criterion. *** indicates that the parameter is significantly different from zero at a 1% level of confidence, ** at 5% and * at 10%. The GMNL model has $\gamma = 0$.

Table VI: Heteroskedastic rank-ordered GMNL and MXL on reduced sample. Standard errors in parentheses.

| | HROGMNL | | MXL - RS | |
	Mean	St.Dev.	Mean	St.Dev.
	Models			
Salary	2.999***	3.360***	3.100***	3.054***
	(0.321)	(0.354)	(0.374)	(0.451)
Supp mgt	1.993***	1.482***	2.040***	1.509***
	(0.198)	(0.159)	(0.236)	(0.218)
Excell care	1.621***	1.400***	1.485***	1.274***
	(0.162)	(0.145)	(0.180)	(0.157)
App resp	0.972***	0.845***	0.952***	0.870***
	(0.111)	(0.128)	(0.148)	(0.210)
Flex rost	0.941***	1.038***	0.926***	1.013***
	(0.112)	(0.125)	(0.137)	(0.186)
Encourage	0.935***	0.788***	0.861***	0.613***
	(0.103)	(0.115)	(0.118)	(0.212)
Well equip	0.785***	0.506***	0.725***	0.707***
	(0.093)	(0.133)	(0.129)	(0.201)
Well staff	0.819***	0.722***	0.755***	0.630***
	(0.094)	(0.107)	(0.115)	(0.183)
Public hosp	0.307***	0.450***	0.549***	0.738***
	(0.064)	(0.154)	(0.115)	(0.229)
3 rotations	0.371***	0.511***	0.389***	0.674***
	(0.069)	(0.115)	(0.105)	(0.192)
Flex hours	0.207***	0.672***	0.212**	0.675***
	(0.063)	(0.107)	(0.091)	(0.207)
Parking	0.159***	0.476***	0.105	0.714***
	(0.061)	(0.114)	(0.096)	(0.191)
Job B Cst	0.118**	0.406***	0.246*	0.069
	(0.052)	(0.075)	(0.136)	(0.098)
Job A Cst	0.012	0.227**	0.260*	0.548***
	(0.048)	(0.112)	(0.146)	(0.173)
δ	-0.021			
	(0.100)			
τ	0.712***			
	(0.084)			
Sample Size	21040		6768	
SLLikelihood	-5751.731		-1761.596	
AIC	11563.462		3579.193	
BIC	11802.087		3770.152	

Notes: HROGMNL refers to a heteroskedastic rank ordered generalised mixed logit model, MXL-RS refers to a mixed logit estimated on the reduced sample of those respondents who completed the survey within 10 days of receiving the link to the website. For the simulations, 10,000 Halton draws are made after burning the initial 43 draws. The coefficient on salary measures the change in utility caused by moving from a job with a weekly salary of 800 to a job with a weekly salary of 1250. For the MXL the standard errors are robust to arbitrary heteroskedasticity and to correlations across observations from the same individuals. SLLikelihood indicates a simulated log likelihood, AIC refers to the Akaike information criterion and BIC to the Bayesian information criterion. For the rank ordered model, BIC is calculated using a ranking as an observation. *** indicates that the parameter is significantly different from zero at a 1% level of confidence, ** at 5% and * at 10%. The HROGMNL model has $\gamma = 0$.

Table VII: Predicted probabilities of job choice by attribute, various models.

	MNL	MXL	GMNL	HROGMNL	MXL-RS
Salary	0.774	0.908	0.968	0.915	0.921
Supp mgt	0.740	0.875	0.946	0.880	0.885
Excell care	0.697	0.808	0.891	0.835	0.815
App resp	0.617	0.723	0.796	0.726	0.722
Flex rost	0.632	0.713	0.796	0.719	0.716
Encourage	0.627	0.695	0.778	0.718	0.703
Well equip	0.592	0.671	0.742	0.687	0.674
Well staff	0.599	0.664	0.741	0.694	0.680
Public hosp	0.560	0.608	0.650	0.576	0.634
3 rotations	0.551	0.593	0.633	0.592	0.596
Flex hours	0.532	0.552	0.580	0.551	0.553
Parking	0.516	0.525	0.540	0.540	0.526

Notes: Figures measure predicted probabilities of job choice (relative to the base job) when the attribute is set to its preferred level, all other attributes remaining at their base level. The base job is one with all attributes set at their least preferred levels. For the salary the shift is from 800 to 1250 dollars per week, for all other attributes the shift is from zero to one. MNL refers to a multinomial logit, MXL to a mixed logit, GMNL to a generalised multinomial logit, HROGMNL to a heteroskedastic rank ordered generalised multinomial logit and MXL-RS to a mixed logit estimated on the reduced sample of those respondents who completed the survey within 10 days. All predicted probabilities are significantly different from 0.5 at the 1% level except for those corresponding to "Parking"; for the latter only the probability in the HROGMNL is significantly different to one half at 1%.

Table VIII: Willingness-to-pay for job attributes, various models.

	MNL	MXL	GMNL	HROGMNL	MXL-RS
a) Coefficients set at their means in the random coefficient models:					
Supp mgt	252.304	252.719	251.256	249.467	247.531
Excell care	208.439	195.663	192.950	209.697	188.982
App resp	126.680	136.807	131.200	133.374	126.949
Flex rost	142.919	130.379	130.849	129.435	123.775
Encourage	137.379	118.585	121.614	128.718	115.756
Well equip	101.530	103.848	103.594	109.576	98.671
Well staff	108.105	99.815	103.265	113.898	102.427
Public hosp	67.058	65.923	62.443	44.828	75.928
3 rotations	57.391	56.393	55.221	53.774	54.582
Flex hours	36.346	32.124	33.095	30.411	30.163
Parking	18.400	15.604	16.546	23.553	15.129
b) Median of the simulated WTP distribution (first and third quartiles in parentheses):					
Supp mgt		175.647	180.457	153.665	169.734
		(41,330)	(51,333)	(9,307)	(34,325)
Excell care		130.674	132.899	125.839	126.635
		(1,273)	(5,274)	(-12,269)	(9,261)
App resp		89.301	88.266	77.349	82.720
		(-14,203)	(-12,199)	(-9,173)	(0,180)
Flex rost		86.383	88.666	73.883	79.688
		(-1,188)	(2,189)	(-29,181)	(-16,185)
Encourage		78.763	83.044	74.303	76.884
		(14,162)	(18,169)	(-6,166)	(16,157)
Well equip		68.457	70.009	65.403	63.506
		(4,148)	(7,150)	(10,136)	(-4,145)
Well staff		66.095	69.874	65.715	67.176
		(8,140)	(10,147)	(-8,150)	(6,145)
Public hosp		40.976	40.280	24.355	47.297
		(-35,122)	(-26,113)	(-19,72)	(-22,124)
3 rotations		34.307	34.480	29.048	33.099
		(-47,118)	(-41,114)	(-21,84)	(-31,102)
Flex hours		19.141	20.934	16.035	17.704
		(-40,80)	(-34,78)	(-48,80)	(-47,82)
Parking		9.598	10.522	12.415	9.431
		(-34,54)	(-37,59)	(-32,59)	(-59,77)

Notes: Figures represent marginal rates of substitution (in absolute value) between the attributes and salary and should be compared to a base salary of $800 per week. In the MNL, point estimates are used to evaluate the MRS. In models with individual heterogeneity, two estimates are provided. In the top panel, WTP is measured with the coefficients set at their mean values. In panel b, the distribution of the WTP measure is simulated with 100,000 replications.

Figure 2: Quantiles of the willingness-to-pay distribution
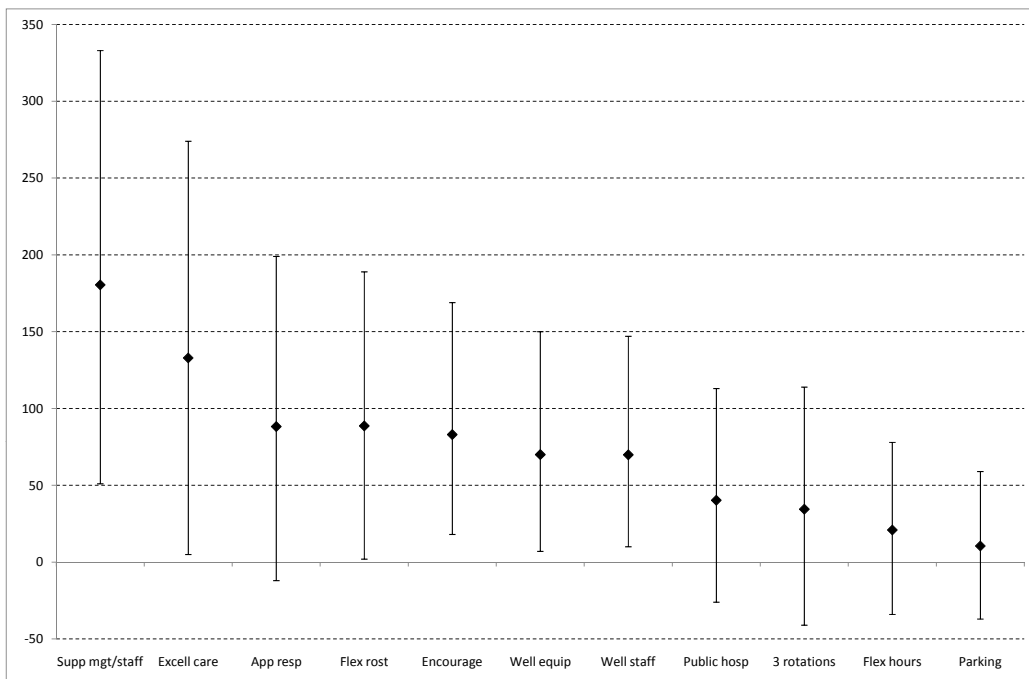
Notes: For each attribute, the median and the interquartile range of the simulated WTP distributions are shown. WTP figures represent marginal rates of substitution (in absolute value) between the attributes and salary and should be compared to a base salary of $800 per week. The distribution of the WTP measure is simulated with 100,000 replications.

Table IX: Mixed logit with shifts in mean attribute weights by year in program. Standard errors in parentheses.

	Mean				SD
	1st yr	2nd yr	3rd yr	Grad	
Log(salary)	4.731***	1.739**	1.047	0.052	5.222***
	(0.642)	(0.836)	(0.832)	(0.981)	(0.605)
Supp mgt	1.909***	0.414*	−0.055	0.308	1.436***
	(0.203)	(0.248)	(0.241)	(0.288)	(0.163)
Excell care	1.788***	−0.314	−0.512**	−0.730***	1.362***
	(0.186)	(0.235)	(0.220)	(0.283)	(0.136)
App resp	1.100***	−0.096	−0.364*	0.099	1.049***
	(0.167)	(0.224)	(0.214)	(0.256)	(0.154)
Flex rost	0.652***	0.535***	0.577***	0.130	0.883***
	(0.130)	(0.198)	(0.195)	(0.221)	(0.142)
Encourage	0.850***	0.054	0.092	−0.074	0.722***
	(0.140)	(0.183)	(0.195)	(0.236)	(0.152)
Well equip	0.671***	0.375*	−0.016	−0.178	0.633***
	(0.128)	(0.201)	(0.179)	(0.231)	(0.161)
Well staff	0.647***	0.038	0.124	0.283	0.603***
	(0.117)	(0.164)	(0.165)	(0.231)	(0.142)
Public hosp	0.401***	0.012	0.105	0.165	0.766***
	(0.133)	(0.198)	(0.190)	(0.223)	(0.155)
3 rotations	0.075	0.326*	0.483**	0.747***	0.821***
	(0.113)	(0.197)	(0.197)	(0.253)	(0.144)
Flex hours	0.097	0.080	0.211	0.434**	0.634***
	(0.105)	(0.155)	(0.175)	(0.207)	(0.153)
Parking	0.204*	−0.234	−0.077	−0.120	0.511***
	(0.112)	(0.175)	(0.170)	(0.203)	(0.164)
Job B Cst	0.354***	0.294*			
	(0.099)	(0.168)			
Job A Cst	0.245**	0.392**			
	(0.100)	(0.160)			
Sample Size	12624				
Log likelihood	−3252.115				
AIC	6632.230				
BIC	7108.605				

Table X: Predicted probabilities of job choice and willingness-to-pay for job attributes, variation by year in program.

| | Rank | Value | Differences from year 1 | | | Rank |
	Year 1	Year 1	Year 2	Year 3	Graduate	Graduate
a) Predicted probabilities:						
Salary	1	0.892†††	0.055	0.037	0.002	2
Supp mgt	2	0.871†††	0.04	−0.006	0.031	1
Excell care	3	0.857†††	−0.043	−0.075**	−0.114**	4
App resp	4	0.750†††	−0.018	−0.074*	0.018	3
Flex rost	7	0.657†††	0.109***	0.116***	0.029	7
Encourage	5	0.700†††	0.011	0.019	−0.016	8
Well equip	6	0.662†††	0.078*	−0.004	−0.041	11
Well staff	8	0.656†††	0.008	0.027	0.061	5
Public hosp	9	0.599†††	0.003	0.025	0.039	9
3 rotations	12	0.519††	0.080*	0.117**	0.176***	6
Flex hours	11	0.524	0.020	0.052	0.105**	10
Parking	10	0.551††	−0.058	−0.019	−0.030	12
b) Willingness-to-pay:						
Supp mgt	1	265.635***	−24.321	−46.064	31.120	1
Excell care	2	251.806***	−88.836**	−93.215**	−93.016*	3
App resp	3	165.962***	−50.986	−70.252**	11.435	2
Flex rost	6	102.967***	31.079	50.321*	17.725	6
Encourage	4	131.513***	−27.225	−11.230	−11.712	7
Well equip	5	105.792***	13.653	−20.045	−27.493	10
Well staff	7	102.196***	−21.904	−2.285	39.073	4
Public hosp	8	65.037***	−15.579	2.015	24.330	8
3 rotations	11	12.659	35.518	61.096**	113.731***	5
Flex hours	10	16.244	5.395	25.318	67.830**	9
Parking	9	33.812*	−37.436	−16.373	−19.844	11

Notes: The Value year 1 column shows the predicted probabilities and WTP figures for year one nursing students. Years 2 and 3 and graduate show shifts in year 1 mean attributes. Rank year 1 and rank graduate show the rankings of the probabilities and WTP figures for year 1 students and graduates respectively. WTP measures are evaluated at the mean attribute levels. *** indicates that the parameter is significantly different from zero at a 1% level of confidence, ** at 5% and * at 10%. Similarly †††indicates that the parameter is significantly different from 0.5 at a 1% level of confidence, ††at 5% and †at 10%. Underlying standard errors are computed using the delta method.