Handling resolvable uncertainty from incomplete scenarios in future doctors' job choice – probabilities vs discrete choices

Accepted for publication in *The Journal of Choice Modelling* 19/11/2019

Line Bjørnskov Pedersen (Corresponding author)
DaCHE –Danish Centre for Health Economics, Department of Public Health, University of Southern Denmark, J.B. Winsløwsvej 9B, 5000 Odense C, Denmark
Research Unit for General Practice, University of Southern Denmark, J.B. Winsløwsvej 9A, 5000 Odense C
email: lib@sdu.dk

Morten Raun Mørkbak
Incentive, Holte Stationsvej 14,1, 2840 Holte, Denmark

Riccardo Scarpa
Durham University Business School, Durham University, Mill Hill Lane, Durham DH1 3LB, United Kingdom
Department of Business Administration, University of Verona, Via Cantarane, 24 - 37129 Verona, Italy
School of Accounting, Economics & Finance, University of Waikato, Te Whare Wananga o Waikato, Gate 1 Knighton Road, Private Bag 3105, Hamilton 3240, New Zealand

ABSTRACT

Health economists often use discrete choice experiments (DCEs) to predict behavior, as actual market data is often unavailable. Manski (1990) argues that due to the incompleteness of the hypothetical scenarios used in DCEs, substantial uncertainty surrounds stated choice. Uncertainty can be decomposed into "resolvable" and "unresolvable"; the former is expected to become resolved in actual choice, as individuals collect further information. To enable its identification, Manski suggests eliciting subjective choice probabilities (ECPs) rather than discrete choices. We introduce the ECP approach in health economics and explore its convergent validity. The context is future physicians' stated choices of job in rural general practice in Denmark. Our results are mixed, but show remarkable similarities in forecasting abilities, despite the ECP models being less econometrically demanding and relying on milder preference distributional assumptions.

Keywords: Discrete choice experiments; elicited choice probabilities; resolvable uncertainty; quantile regression; rural general practice

JEL classification: C31; C35; I11

1. Introduction

Forecasting choice behavior in health economics is challenging because actual data is often unavailable. In order to derive estimates of health care demand and/or supply, health economists often resort to data derived from hypothetical choice scenarios. An increasingly popular way of doing this is by means of discrete choice experiments (DCEs) in which respondents select their favorite alternative between two or more hypothetical scenarios describing the available treatments, services or, in this case, general practices. Data collected in this manner are then used to estimate random utility models, based on specific assumptions on behavior, from which expectations of real choice behavior are forecast (see e.g. Brown et al. 2015, Meenakshi et al. 2012, Sivey et al. 2012, Lancsar et al. 2011). Manski (1990) argues that statements of choice intentions might not be good predictors of future behavior. This because respondents taking part in a hypothetical DCE are likely to be provided with only a subset of the information deemed subjectively relevant or even necessary to make a real-life choice. Manski (1999) referred to this limitation as DCEs suffering from "incomplete scenarios" and repeated what Fischhoff, Welch, and Frederick (1999) wrote "If needed detail is missing, then people may make it up."

Uncertainty generates a systematic divergence between hypothetical and actual choice. The divergence is caused by the necessarily incomplete information base provided to respondents when eliciting their choice intentions. This information gap causes what was termed "resolvable uncertainty" (Manski 1999) where the term "resolvable" indicates that to make real life choices respondents expect (and are expected) to acquire further information in order to reduce the overall uncertainty. In our study, we deal with the preferences future physicians hold over types of employment as general practitioners (GPs) in a general practice in rural locations in Denmark. This is currently an important area of investigation as there is a severe shortage of GPs (Marchand and Peckham 2017). Also, it is likely that when junior doctors make actual job choices, they would further seek and process details about working conditions in different general practices. Hence, it is plausible to expect that once faced with a real choice scenario, respondents would have resolved some of the uncertainty surrounding the hypothetical choice context. Cognizant of this fact, analysts are faced by an extrapolation problem in which assumptions are likely to play a crucial role. However, the consequences of such information discrepancy and their impact on standard assumptions in the analysis of choice data have rarely been explored in the empirical literature.

In the standard DCE framework, the issue of incomplete scenarios is typically handled by assuming that what remains undescribed in the characterization of alternatives has no systematic effect on utilities of alternatives as evaluated by respondents, except perhaps by inflating error variance. Some studies report asking respondents to score how certain they are about their stated choice and they use this score in heteroskedastic choice models. Because of limitations in both cognitive effort in information processing and in experimental design dimensions, it is of course impossible to include all potentially salient characteristics of alternatives in a DCE setting. However, eliciting subjective choice probabilities could potentially overcome this issue, by allowing respondents to be explicitly uncertain about their stated subjective choice. This approach was first proposed by Manski (1999) and later applied by Blass et al. (2010) who also demonstrated its additional advantage in giving rise to specifications with estimations that are substantively less econometrically demanding, thereby giving rise to a claim of a more robust inference.

The elicitation of subjective choice probabilities has only been applied in a few stated choice studies, and never in the area of health economics. Blass et al. (2010) show how the approach can empirically be applied to data on consumers' preferences for the reliability of electricity services in Israel. Shoyama et al. (2013) used this approach for eliciting public preference for land-use scenarios in Kushiro watershed in northern Japan and find some divergence between willingness to pay (WTP) estimates obtained from standard DCE data and the alternative elicited choice probability (ECP) approach. In a working paper on lake recreation, Herriges, Bhattacharjee and Kling (2011) use the 2009 Iowa Lake Survey to administer a split treatment in terms of information provision (low and high) and preference elicitation method (DCE versus ECP). They find significant differences between the two formats in terms of implied preferences for two hypothetical lake scenarios. More generally, subjective choice probabilities have recently been used within labor economics – studying e.g. the choices of major subjects by college students and income expectations in American households, although not in a choice experiment framework (Dominitz and Manski 1997, Arcidiacono et al. 2012; Wiswall and Zafar 2015).

In this study we contribute to the sparse literature on this issue by examining the convergent validity of the DCE and the ECP approaches. We focus on the hypothesis that, when some resolvable uncertainty is allowed to be expressed by using subjective probabilities, the structure of utility differ from that underlying discrete choices where this uncertainty cannot be expressed.

Unlike previous studies using this approach, we sample a population of future medical doctors. Compared to the populations sampled in other studies, our respondents are extremely well-educated and have all been exposed to academic training in the field of probabilities. We first analyze the data from the DCE using conventional approaches based on random parameter mixed logit (RPL) models of discrete choice. Then from the ECP using both RPL models and, following Blass et al (2010) the more robust least absolute deviations (LAD) regressions.

2. Methods

2.1. Stated discrete choices

The underlying theory of DCE is based on Lancaster's consumer theory (Lancaster 1966) and random utility theory (Luce 1959; McFadden 1974). The model specification most commonly applied in the parametric analysis of preferred choice responses is currently the mixed logit model. This is based on the random utility one can derive from each alternative *i* offered in the choice set (see Train 2009; Hensher and Greene 2003):

$$U_{ntk} = V(x_{nti}, \beta_n) + \varepsilon_{nti}. \quad (1)$$

The individuals are denoted by *n,* while *i* denotes the alternative and *t* is the choice task in the sequence of *T* observed choices *t*=1,2,…,*T*. The indirect utility, *V*, is a function of the vector of attributes describing the attractiveness of jobs, x_{nti}, and the associated vector of individual taste parameters, β_n. The utility component ε_{nti} is unobserved by the researcher and hence assumed to be i.i.d. extreme-value distributed. The random utility specification implies that only part of the deterministic (indirect) utility function is known by the analyst, while the respondent/individual is fully aware of the exact level of utility (s)he will obtain from a given alternative. Following Train (2009), the probabilities of the mixed logit model can be described as integrals of the standard conditional logit function evaluated at different values of β_n distributed according to a density function (the mixing distribution). This specification can be generalized to allow for a panel of *t*=1,…,*T* choices by the same respondent n. Thus, the respondent-specific vector of taste parameters may vary over the population of future physicians but are constant over the *T* choice occasions by the same physician. The joint unconditional choice probability for the sequence of *T* choices by respondent *n* in the sample, under the mixed logit assumption, becomes:

$$P_n = \int \left(\prod_{t=1}^{T} \left[\frac{e^{\lambda \beta_n{}' x_{ntk}}}{\sum_i^J e^{\lambda \beta_n{}' x_{nti}}} \right] \right) f(\beta|\mu, \eta) d\beta \quad (2)$$

where f is the density distribution function for β_n with a mean of μ and a variance-covariance of η, and λ is the scale parameter, which is typically normalized to unity to allow identification.

In our case, all coefficient attributes describing the job option are assumed to follow a normal distribution, since these can be seen as desirable or undesirable, depending on respondents' preferences. The coefficient for the income attribute, instead, is assumed to be fixed. For further detail of the RPL model see e.g. Train (2009).

We also estimate the WTP estimates by taking the ratio of the estimated mean parameters. By keeping the income coefficient fixed, we avoid the issue of potentially undefined distributions of the WTP obtained from taking the ratio of two random coefficients with different underlying distributions – e.g. a normal and a log-normal distribution. Estimation of WTP enables us to compare the size of the estimates in the different models, and to avoid the issue of scale effects discussed in Swait and Louviere (1993). By using symmetric preference distributions and a fixed cost/income coefficient, the estimated mean WTPs are equivalent across mixed logit and LAD models.

2.2. Elicited choice probabilities

As noted above, stated choice analysis assumes that respondents know the value of the utility of each alternative—i.e. both V and ε—which make them able to identify the utility maximizing alternative in all stated choice tasks. As argued in the introduction, and as eloquently put forth by Manski (1999), this is rarely the case in a hypothetical scenario used to describe experimental choice situations. It stands to reason that there are descriptors of alternatives that respondents would find informative in real life choices but are left totally omitted or insufficiently described in the survey scenarios. When approaching real life decisions some of these would become known to respondents—our future physicians would expect to "resolve" at least some of the uncertainty present in the hypothetical scenarios of the survey when choosing a real job in general practice. They would also hold some subjective belief with regards to these "resolvable features", with distributions that correlates with scenarios' descriptors. So, following the random utility framework, but extending it to account for the uncertainty created by the incomplete alternative descriptions salient for individual utility, we can rewrite the utility function in (1) as:

$$U_{nti} = V(x_{nti}, \beta_n) + \varepsilon_{nti} = V(x_{nti}, \beta) + \varepsilon_{nti}^{r} + \varepsilon_{nti}^{u} \quad (3)$$

where ε_{nti}^{u} includes unresolvable uncertainty, which is treated as in the conventional random utility specification, instead, ε_{nti}^{r} is a factor related to idiosyncratic *resolvable* uncertainty. We assume here, for simplicity, that this is the only component of idiosyncratic randomness of utility coefficients, but it need not be so. The problem in stated preference experiments is that analysts will never come to the ex-post real choice situation as choices remains hypothetical at the moment of data analysis. This is why it seems important to know how to treat this type of resolvable uncertainty separately into the stated preference experiment format.

Since ε_{nti}^{r} is unknown to the decision maker at the moment of expressing the stated choice, we can no longer assume that (s)he has made a definite utility-maximizing selection when choosing between hypothetical alternatives. Instead one can more plausibly say that (s)he has made a selection according to a subjective assessment concerning which alternative will yield the largest utility (allowing for uncertainty across alternatives). Now—following Blass et al. (2010)—this can still be set up in a random utility framework. Suppose that individual *n* forms a subjective distribution on the values of $\varepsilon_{nti} = \varepsilon_{nti}^{r} + \varepsilon_{nti}^{u}$, which makes him/her capable of choosing alternative *i* at time *t* with a given subjective probability q_{nti}. Assume that individual *n* has the utility function given by eq. 3, and given the attributes in *V,* as represented in the vector x_{ntk}, places a subjective distribution Q_{nt} on ε_{nti}, then we can write the subjective choice probability for individual *n* choosing alternative *i* over *j* in choice situation *t* as:

$$q_{nti} = Q_{nt}\left[x_{nti}\beta_n + \varepsilon_{nti}^{u} > x_{ntj}\beta_n + \varepsilon_{ntj}^{u},\ \ for\ all\ i \neq j\right] \qquad (4)$$

The subjective distribution Q_{nt} reflects the resolvable uncertainty in a hypothetical stated choice situation, which is unknown at that stage, but will become known in the actual choice situation. Hence, to point out the salient difference between stated choice (which only allows for unresolvable uncertainty) and elicited choice probabilities (that allows for both resolvable and unresolvable uncertainty) imagine that when respondent *n* states that (s)he will choose alternative *i* over *j* – in the elicited choice based on subjective probability format, this means that $q_{nti} \geq q_{ntj}$. This inequality does not necessarily mean, as instead is assumed in stated choice surveys, that $U_{nti} \geq U_{ntj}$. The discrepancy is to be found in the subjective features of resolvable uncertainty. Therefore, only when $q_{nti} = 1$ is the subjective probability format in accordance with the assumption in stated choice analysis. This difference is more salient the closer resolvable uncertainty makes q_{nti} proximate to 0.5. For example, in a sample with 100 choice tasks in which $q_{nti} = 0.6$ will lead to only 60 of these selecting alternative i, while in the stated choice analysis all 100 of them will be taken as selecting alternative i. Non-extreme subjective probabilities indicate belief of resolvable uncertainty, which respondents expect to solve in an actual choice.

The next step is to estimate the subjective random utility model, which requires some standard assumptions analog to the assumptions made in the standard stated choice analysis. Now – assume that the ε_{nti}^{u} is *subjectively* i.i.d. distributed Gumbel (as opposed to being *objectively* i.i.d. distributed in the standard model), which then gives us the logit choice probability:

$$q_{nti} = \frac{\exp(\lambda\beta_n' x_{nti})}{\sum_j^J \exp(\lambda\beta_n' x_{ntj})},\ j,i=1,\ldots,J. \tag{5}$$

In a binary choice context, one has:

$$q_{nt1} = \frac{\exp(\lambda\beta_n' x_{nt1})}{\exp(\lambda\beta_n' x_{nt1})+\exp(\lambda\beta_n' x_{nt2})} \tag{6}$$

and

$$q_{nt2} = \frac{\exp(\lambda\beta_n' x_{nt2})}{\exp(\lambda\beta_n' x_{nt1})+\exp(\lambda\beta_n' x_{nt2})} \tag{7}$$

which leads to:

$$\ln\left(\frac{q_{nt2}}{q_{nt1}}\right) = \ln\left(\frac{\exp(\lambda\beta_n' x_{nt2})}{\exp(\lambda\beta_n' x_{nt1})}\right) = \beta_n{}'(x_{nt2} - x_{nt1}) \tag{8}$$

So, simply regressing the log-odds on the internal product of coefficients and attribute differences leads to a linear probability model.

Assume now that the single attribute coefficient is individual-specific because of an idiosyncratic component ϵ_n, then $\beta_n = b + \epsilon_n$:

$$\ln\left(\frac{q_{nt2}}{q_{nt1}}\right) = (x_{nt2} - x_{nt1})\beta_n = (x_{nt2} - x_{nt1})(b + \epsilon_n) = (x_{nt2} - x_{nt1})b + (x_{nt2} - x_{nt1})\epsilon_n, \tag{9}$$

Setting $u_{n21} = (x_{n2} - x_{n1})\epsilon_{n21}$ and ignoring *t,* the above collapses to:

$$\ln\left(\frac{q_{n2}}{q_{n1}}\right) = (x_{n2} - x_{n1})\beta_n = (x_{n2} - x_{n1})(b + \epsilon_n) = (x_{n2} - x_{n1})b + u_{n21}, \tag{10}$$

Without loss of generality, setting $E(\epsilon_n|x) = 0$ by normalization, and taking expectations, one finds $E(u_n|x) = 0 \; and \; b = E(\beta)$ thus making eq. (10) a linear mean regression model:

$$E\left[\ln\left(\frac{q_{n2}}{q_{n1}}\right)|x\right] = (x_{ntk} - x_{nt1})b \tag{11}$$

The above relationship implies that the mean value of the preference in this binary ECP model can be consistently estimated by OLS with only week distributional assumptions on the random $\beta_n = (b + \epsilon_n)$, which only require a null value of the conditional expectation. This is a much milder assumption than what is normally assumed in mixed logit with normally distributed coefficients, in which the required assumption is $\epsilon_n \sim \varphi(0, \sigma)$. Note that in the latter a very specific type of symmetry is required around zero, determined by both the size of σ and the shape of the normal. No assumption about shape or σ is required in the ECP.

Unfortunately, equation (11) may not be taken at face value under all empirical circumstances. A typical problem arises in the frequent instances in which respondents round elicited probabilities to the closest 5 or 10% (Manski, 2004; Manski and Molinari 2010). The problem is immaterial when the rounding takes place well within the (0,1) interval, but it becomes numerically serious if small subjective probabilities are rounded to 0 and large ones are rounded to 1, as the log-odds are very sensitive near the boundaries (0 and 1): in the extreme case they will end up providing log-odds of either plus or minus infinity, making OLS estimates computationally infeasible. As shown by Blass et al. (2010) this problem is solved if the preferences are symmetrically distributed around their median value, as is also assumed in many stated choice analyses using normal or equilateral triangular distributions of the random parameters. Such symmetry implies that the unobservable idiosyncratic component of the random coefficient vector $\Delta x'\epsilon_n = u_n$ in eq. (10) be symmetrically distributed with zero median, conditional on x. The linear median regression model is robust to extreme values and can be consistently estimated by using Least Absolute Deviations (LAD) (Koenker and Bassett, 1978; Koenker and Hallock, 2001):

$$M\left[\ln\left(\frac{q_{nt2}}{q_{nt1}}\right)|x\right] = (x_{nt2} - x_{nt1})b \tag{12}$$

We take advantage of the property of the median (indeed any quantile) of a random variable being invariant towards any order-preserving transformation. This enables us to substitute reported values of zero and one for values close to them, but in the interior of the (0,1) interval[1], without affecting the consistency of the estimate of the mean value for the random coefficient, or b in eq. (11).

Because the LAD approach provides consistent and unbiased coefficient estimates, the derivation of marginal WTP for choice attributes at the mean/median values of the random coefficients is also straightforward and equals the (negative) ratio between the mean/median of any random coefficient for a given job characteristic and the mean/median of the random coefficient for the pecuniary attribute. This is resting entirely on symmetry assumptions similar to those invoked in the mixed logit with normally distributed random coefficients and fixed price coefficient, which is still very commonly employed in practice (see the review in Caputo et al. 2018). But it enables the estimation of the parameters of interest from probability data instead of choice data, without invoking parametric distributional assumptions beyond linearity. We also note that while the mean values may not exist in the sample or population, quantile values, such as the median, always exist. What may not exist is a respondent that holds the median value of the random coefficient at the numerator and the median value of that at the denominator *at the same time*, but this is a shortcoming that can also be attributed to mixed logit with normal random parameters for which this WTP measure is often used.

2.3. A comparison of preference elicitation approaches: DCE vs ECP

We compare the data obtained from the two independent samples treated with DCE and ECP elicitation formats in a number of different ways. After having tested for successful randomization of respondents between treatment groups using *Chi*-squared tests, we make descriptive comparisons of choices, and compare the perceived certainty of answers, measured in a follow-up question in the two treatment groups, where for "treatment" we intend the preference elicitation format: DCE and ECP. In the ECP treatment group, we also compare the degree of certainty by contrasting respondents who *always* stated probabilities of 0% and 100% with those who expressed *at least some* interior probabilities. We then compare the signs and significance of mean/median estimates across models. Note that RPL specifications can be estimated from both DCE and ECP[2] data, while LAD models only from ECP data. We then test for differences in WTPs and for taste heterogeneity across treatments and

[1] To enable the log-odd computation, in the execution of the analysis of the data we transform reported ones and zeros to 0.999 and 0.001, respectively. This does not affect the LAD estimates, which is invariant to truncation of extreme values. Note that in the text we follow Blass (2010) and refer to m as the "mean preferences" rather than the more correct "center of symmetry of the preference distribution".

[2] The ECP choice data are recoded as one for alternatives with stated probability greater than 50%, and zero for stated probabilities below 50%. Choice tasks with subjective probabilities of 50% are excluded in this analysis.

models. Finally, we examine the predictive validity of the DCE and the ECP approaches by evaluating the forecasts of each of the models.

In the DCE treatment, we forecast choice probabilities for all choice tasks and compare these with the actual choices using a threshold of 0.5, which is adequate for binary choices. The ECP forecast, instead, is deemed correct if both forecast and the stated probability assign more than 0.5 to the same alternative[3]. We report the percentage of correct predictions for both treatments. We perform the convergent validity test in three different ways:

1) A within-respondent consistency test by comparing predicted choice probabilities with stated choices (using the RPL model in the DCE sub-sample[4] and the LAD model in the ECP sub-sample).
2) A between-respondent consistency test *across* treatment groups by comparing probability forecast from the RPL model onto the ECP responses and probability forecast from the LAD for the DCE responses[5].
3) A between-respondent consistency test *within* each treatment group by comparing predicted choice probabilities obtained from a model estimated on a randomly selected subset of 85% of respondents with a hold-out subsample of actual choices/choice probabilities of the remaining 15% of respondents.[6]

3. Case study: Future physicians' preferences for rural general practice

Like many other countries, Denmark has been suffering from a shortage of general practitioners (GPs). This shortage is especially felt in rural areas, and the issue is unlikely to be alleviated in years to come. Designing policies to attract physicians (both future and existing) to rural areas has long been a crucial challenge for health economists and others, who have addressed it in the literature a number of times (Marchand and Peckham 2017). To shed light on the issue, previous studies used both choice experiments (see e.g. Scanlan et al. (2018), Holte et al. (2015) and Li et al. (2014)) and best-worst scaling (Günther et al. 2010). To date, the ECP approach has never been used in a health economics framework.

[3] The exact 50% responses have been removed for the purpose of this test.
[4] This approach is also used in Salampessy et al. (2015).
[5] If we find major differences in preference structure between the two approaches this test cannot be seen as a validity test.
[6] This approach is similar to the one used by Salampessy et al. (2015) except that they use 50% of the sample to make predictions of the choices made by the remaining 50% in a DCE setting.

3.1. Survey design and questionnaire

The first section of the questionnaire included some preliminary questions about respondents' place of study, preferred choice of medical specialty and general reasons for wanting to be a GP. The second section included the discrete choice/subjective choice probability experiment. This was followed by a question on perceived certainty in choices together with socio-economic questions. The questionnaire was pilot-tested on 21 medical students. No major changes were made following the pilot.

Respondents are future medical doctors (medical students in their final years of study) and each was assigned 12 experimentally designed choice scenarios consisting of two alternatives ($J = 2$). Allocation to preference elicitation treatment was randomized based on their date of birth: even dates were allocated to the ECP, odd dates were allocated to the DCE. All other aspects of the experiment, including choice scenarios, were kept identical across treatments.

Scenarios were described by seven attributes (see table A1 in appendix); two with four levels (population and yearly bonus), two with three levels (number of GPs in the practice and distance to closest family), and three with two levels (control over working hours, distance to leisure activities and job security for partner). Attributes and levels were chosen based on both a focus group and extant literature (e.g. Holte et al. 2014; Pedersen and Gyrd-Hansen 2014; Günther et al. 2010; Sivey et al. 2012). Examples of the DCE and ECP treatments are displayed in figure A1 and A2 in appendix.

During the focus group interview, attributes other than those in table 1 were explored and eventually discarded as of minor importance. Dismissed attributes included: professional development opportunities, workload, professional collaboration in the practice, distance to other career options, number of on-call duties, collaboration with other general practices, administrative work, continuity in care, and time for each patient. Resolvable uncertainty surrounding stated choice and pertaining to each of the excluded attributes may, to some extent, be resolved at the stage in which a real job choice is to be made. This suggests that in the present case study, resolvable uncertainty is present.

Although respondents were presented with a forced choice, an 'indifferent option' was available in the DCE treatment. While this is unusual in DCEs practice, we allowed for this option to compare its frequency with the equivalent 50/50 option in the ECP elicitation. In the estimation of choice models from DCE data, observations expressing indifference between alternatives were dropped.

In the ECP treatment, respondents stated their subjective probability of choosing alternative A, the complement of which is the selection probability of alternative B. To help respondents provide us with an accurate probability of a choice, we used a visual aid in the form of a sliding probability scale ranging from 0%-100%. Using the slider bar, respondents could slide back and forth accurately signaling their subjective probability of job selection. The starting point of the slider was at 50-50 percent, and respondents were forced to click on the slider before they could proceed to the next choice set. Previous papers (e.g. Dominitz and Manski 1997, Blass et al. 2010, Shoyama et al. 2013) used open-ended questions, but we feared that this could result into too abstract an exercise for respondents to comprehend. Moreover, Bruine de Bruin et al. (2002) argue that using open-ended questions induces respondents to over-state frequencies around 50%, due to the phrase "fifty-fifty", which to many may not represent a true numeric probability of 50%, but rather some form of epistemic uncertainty (see also Bruine de Bruin et al. 2000, Bruine de Bruin and Fischhoff 2017 and Manski 2017). To avoid this problem, Bruine de Bruin et al. (2002) show that by using an explicit numeric probability scale, as we did here, the issue of exaggerating probabilities of 50% is heavily reduced.

The experimental design was a Bayesian D-error minimizing main effects design with priors estimated from a conditional logit model based on the incoming answers to a pilot test (n=21). The design was generated using the software Ngene (ChoiceMetrics 2009).

3.2. Data collection

Survey data were collected from postgraduate medical students in Denmark in October 2015 using internet forums specifically established for such students at the four universities in Denmark that educate physicians. The link to the questionnaire was distributed in these forums three times during the data collection process. In total, 316 respondents answered the questionnaire, of whom 167 answered the discrete choice questions, and 149 answered the choice probability questions.

4. Results

4.1. Test for successful randomization

The test shows that randomization is successful on all observable characteristics except for respondents from Copenhagen University and Aarhus University (Table A2 in appendix).

4.2. Descriptive comparison of choices

Figure 1 shows that in the DCE treatment, alternative A was chosen in 45.15 % of all choice sets, whereas alternative B was chosen in 46.49% of all choice sets. The indifferent option was chosen in 8.37 % of choices. Removing from the DCE sample those observations stating indifference, caused the distribution to change to 49.27% for alternative A and 50.73% for B.

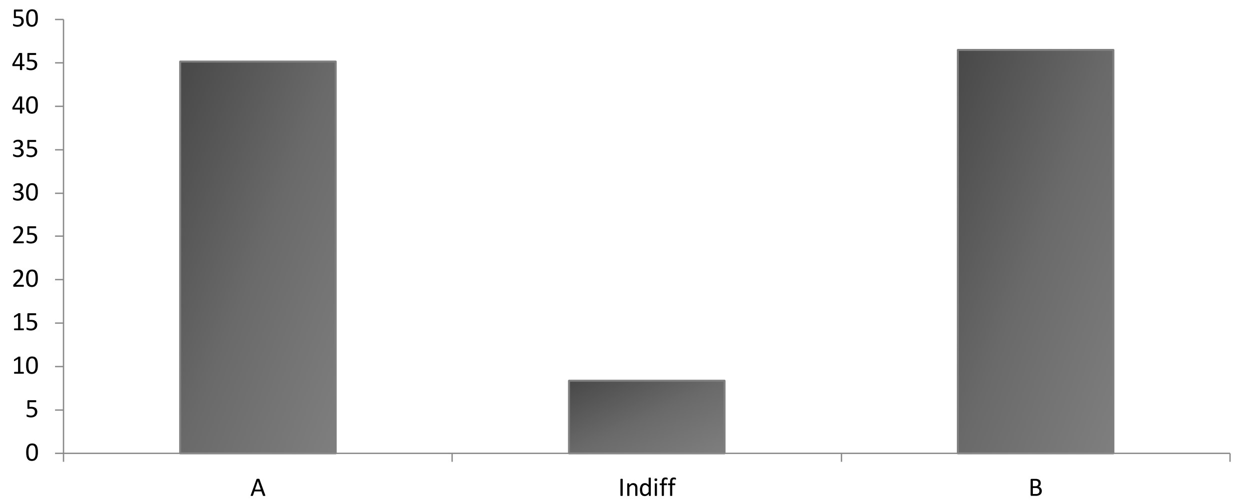

Fig. 1. Distribution of choice frequencies for the DCE treatment including the indifferent option

In the ECP treatment, q_{ntA} and q_{ntB} denote the stated choice probabilities in percent for alternative A and B, respectively, elicited from individual n in a given scenario t. The histogram for the elicited subjective choice probabilities for alternative B are shown in figure 2. For all cases, we have that $q_{ntA} = 100-q_{ntB}$.

The histogram in figure 2 shows that most responses are multiples of 10, followed by multiples of 5, but that other percentages have also been chosen. Disregarding responses at 0% and 100%, the histogram displays a bimodality in choices with some concentration of responses around 20% (chosen 4.39% of times) and 80% (chosen 3.79% of times).

It is important to observe that extreme probabilities of 0% and of 100% have been chosen 7.46 % and 7.68% of the times, respectively. In our proposed interpretation, both cases are equivalent to absence of resolvable uncertainty. So, about 85% of choices are indeed consistent with the simultaneous presence of both resolvable and unresolvable uncertainty, linked to scenario descriptions judged somewhat incomplete to allow for a definitive choice response. Put differently, by allowing an ECP format more information is provided in about 85% of the choices. This wealth of information is obviously unavailable in a conventional DCE, which forces respondents to state only extreme probabilities of 0% and 100%. We note that these figures are comparable to those obtained in Blass et al. (2010), except for the frequency of 50%, which was more seldom selected in our survey. This was perhaps due to our use of the sliding scale, which would be consistent with the findings reported by Bruine de Bruin et al. (2002). We further speculate it could also be due to the inherent comparatively higher ability of our educated respondents to discriminate and interpret the concept of probability.

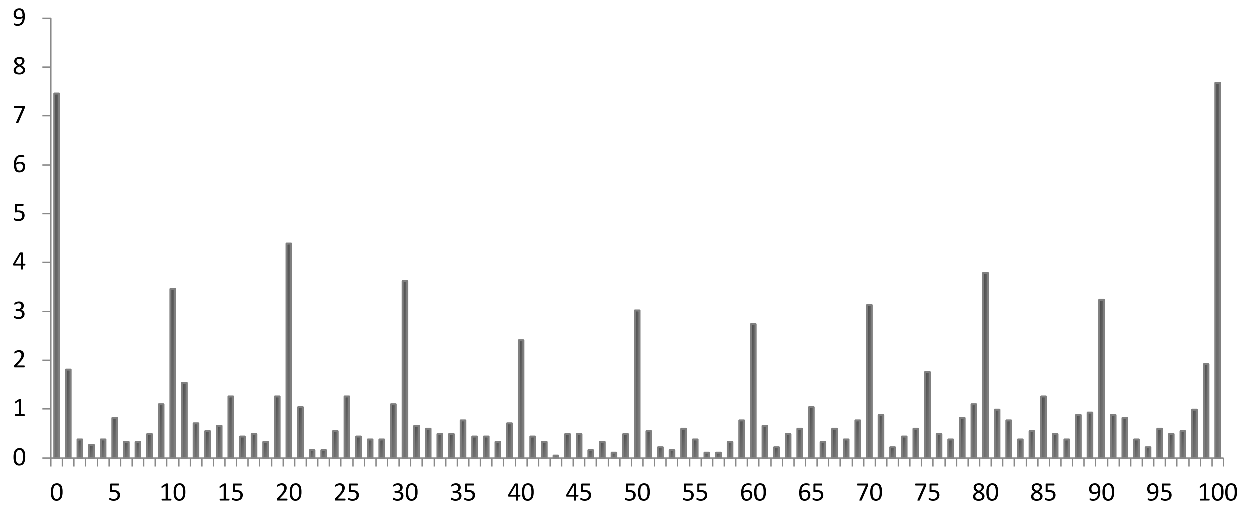

Fig. 2. Distribution of choice frequencies for alternative B in the ECP treatment

For comparison with the DCE treatment, a reported subjective probability below 50% (corresponding to choosing alternative A while allowing for uncertainty) was chosen 47.78% of the times, while a probability above 50% (Corresponding to choosing Alternative B while allowing for uncertainty) was chosen 49.07% of the times. The 50-50 option was chosen only 3.02% of the times.

4.3. Comparison of perceived certainty of choices across treatment groups

Table 1 shows that, regardless of the elicitation treatment, in the follow-up question the majority of respondents state to be uncertain about their stated choice. Respondents in the ECP treatment are slightly more uncertain of their stated probabilities than respondents in the DCE treatment, but the difference is statistically insignificant. Hence, using an ECP rather than a DCE format does not seem to significantly affect the self-reported degree of certainty of responses.

Amongst respondents in the ECP approach who provide interior probabilities, 56.57% said they are uncertain of their response, a percentage that goes up to nearly 69% amongst those only stating limit values (of 0% or 100%).

	DCE	ECP	Pearson Chi2 test	ECP Interior	ECP Zero-one	Pearson Chi2 test
How certain are you of your choices in the last 12 questions?						
Very certain	0.60	0.00		0.00	0.00	
Certain	10.18	7.38		7.71	4.19	
Neither	31.14	28.19		28.99	20.36	
Uncertain	49.10	57.72		56.57	68.86	
Very uncertain	8.98	6.71	0.489	6.72	6.59	0.032

Table 1
Perceived certainty in choices

4.4. Sign and significance of mean attribute preferences

In RPL models, a random coefficient is deemed statistically insignificant if both its mean and standard deviation estimates are insignificant. In table 2 it is seen that for the DCE data this occurs for *familycar, pop2* and *pop25*, suggesting that individuals on average are indifferent between having a short or a long car drive to their closest family relation; or having a practice within sparsely populated rural areas or mid-sized cities. On the contrary, in the set of pseudo-choices derived from the ECP sample (where we used a 50% cut-off), except for the *gp2*, all random coefficient estimates show significance in either the mean, standard deviation, or both. This suggests individuals are on average indifferent between a job in a practice alone or with one other GP. Although the signs of estimated coefficients are all concordant, it appears that the two sub-samples provide partially different patterns of preference heterogeneity. Moreover, the average log-likelihood value for DCE data is 44% higher than for ECP, suggesting a strong inherent difference in response mode.

Table 3 presents estimates for the LAD model fitted to the ECP data. Results reveal no major differences on signs and significance of attribute coefficients, except for the coefficient for *familybike*, which in this data is insignificant, while we would expect it to be positive and significant. No further differences are found when these results are compared with the RPL model fitted to the ECP data (table 2).

4.5. Comparison of marginal WTP estimates

Table 4 shows that WTP estimates generally differ between econometric models (RPL vs. LAD) (WTP estimates appear in table A3 in the Appendix), but not so much between elicitation methods, holding the econometric method constant (RPL(DCE) vs. RPL(ECP)).

4.6. Forecasting

Table 5 shows that for the within-respondent consistency test, the RPL estimates from DCE data (i.e. DCE(RPL)) have a 78% predictive validity (compared to 81% in Salampessy et al. (2015)), while the LAD model has a predictive validity of 75%. Values are reversed in the between-respondent consistency test across treatment groups, as the DCE(RPL) model has 75% predictive validity, and the LAD model 78%. In the between-respondent consistency test within treatment groups (using the hold-out sample), the DCE(RPL) model correctly predicts 79% of choices (compared to 45% in Salampessy et al. (2015)), while the LAD model predicts 75% correctly. For the ECP(RPL) model the value is only 70%. In conclusion, both RPL and LAD models have reasonable predictive performance – in the range of 75-79%. Bearing in mind that the econometrics of the LAD model is predicated on generally weak assumptions: apart from the assumptions necessary for the core logit probability, there is no specific parametric distribution assumed on the random parameters, except median independence and a zero median of the idiosyncratic departure from mean preference values. These assumptions are quite less restrictive, and hence more credible, than those invoked by standard RPL models based on parametric taste distributions. Nevertheless, it would seem that the predictive performance of the LAD model is quite good, and signs of estimates are concordant with those obtained from DCE data with a standard RPL model with normally distributed coefficients.

	DCE				ECP			
	Estimate	St.err.	|*t*-val|	P-value	Value	St.err.	|*t*-val|	P-value
μ ASC2	0.37	0.17	2.16	0.03	2.37	0.34	7.04	<0.001
σ ASC2	*0.01*	*0.23*	*0.06*	*0.95*	*3.74*	*0.42*	*8.99*	*<0.001*
μ bonus1	0.68	0.11	6.41	<0.001	0.70	0.10	7.20	<0.001
σ bonus1	*fixed*				*fixed*			
μ control	2.63	0.50	5.29	<0.001	1.87	0.29	6.51	<0.001
σ control	*1.17*	*0.25*	*4.76*	*<0.001*	*1.59*	*0.26*	*6.04*	*<0.001*
μ familybike	1.99	0.60	3.33	<0.001	0.94	0.35	2.68	0.01
σ familybike	*0.56*	*0.27*	*2.07*	*0.04*	*1.45*	*0.24*	*6.15*	*<0.001*
μ familycar	0.31	0.84	0.37	0.71	0.99	0.40	2.47	0.01
σ familycar	*0.59*	*0.85*	*0.69*	*0.49*	*0.03*	*0.33*	*0.08*	*0.94*
μ gp2	-9.96	4.71	2.12	0.03	-0.25	0.61	0.41	0.68
σ gp2	*6.64*	*2.60*	*2.55*	*0.01*	*1.25*	*0.65*	*1.93*	*0.05*
μ gp34	3.95	1.01	3.91	<0.001	3.47	0.55	6.26	<0.001
σ gp34	*5.98*	*1.66*	*3.60*	*<0.001*	*2.57*	*0.38*	*6.81*	*<0.001*
μ jobhigh	1.60	0.43	3.74	<0.001	2.26	0.39	5.79	<0.001
σ jobhigh	*1.60*	*0.31*	*5.20*	*<0.001*	*2.22*	*0.32*	*6.87*	*<0.001*
μ pop2	-0.50	0.68	0.74	0.46	-1.29	0.41	3.19	<0.001
σ pop2	*0.36*	*0.70*	*0.52*	*0.60*	*0.76*	*0.25*	*3.06*	*<0.001*
μ pop25	-0.47	0.61	0.78	0.44	-1.05	0.39	2.67	0.01
σ pop25	*0.10*	*2.69*	*0.04*	*0.97*	*0.24*	*0.46*	*0.52*	*0.61*
μ pop510	-1.01	0.30	3.31	<0.001	-0.65	0.20	3.35	<0.001
σ pop510	*0.04*	*0.64*	*0.07*	*0.95*	*0.03*	*0.26*	*0.12*	*0.90*
μ schoolbike	2.44	0.39	6.27	<0.001	1.58	0.25	6.45	<0.001
σ schoolbike	*0.83*	*0.29*	*2.82*	*0.01*	*1.28*	*0.20*	*6.28*	*<0.001*
# obs	1851				1822			
LL	-753.44				-1067.04			

Table 2
RPL Models for both DCE and ECP treatment (treated as discrete choice with 3000 Halton draws with Biogeme, Bierlaire 2003)

	LAD		
	Coef.	St.err.	P-value
ASC2/const	0.17	0.05	<0.001
bonus	0.26	0.03	<0.001
control	0.40	0.11	<0.001
familybike	-0.33	0.19	0.08
familycar	1.92	0.19	<0.001
gp2	0.62	0.30	0.04
gp34	1.43	0.27	<0.001
jobhigh	0.37	0.15	0.02
pop2	-2.82	0.27	<0.001
pop25	-2.33	0.26	<0.001
pop510	-0.56	0.07	<0.001
schoolbike	1.22	0.14	<0.001
#obs		1822	
R-squared		0.37	
Log-pseudo-L		N.A.	

Table 3
LAD (50% quantile) for the ECP treatment

	RPL vs. LAD		RPL(DCE) vs. RPL(ECP)	
	t-values	P-value	*t*-values	P-value
control	3.729	<0.001	1.916	0.055
familybike	3.954	<0.001	1.767	0.077
familycar	3.925	<0.001	0.672	0.502
gp2	2.493	0.013	2.103	0.036
gp34	2.044	0.041	0.635	0.525
jobhigh	2.585	0.010	1.012	0.311
pop2	1.765	0.078	0.878	0.380
pop25	5.131	<0.001	0.716	0.474
pop510	4.822	<0.001	1.234	0.217
schoolbike	9.891	<0.001	2.134	0.033

Table 4
Test for differences between WTP estimates using WTP from the RPL model in the DCE treatment as base case

	Correct forecast		
Data(Model)	On DCE data	On ECP data	15% Held out sample
DCE (RPL)	77.90%	74.87%	78.90%
ECP (LAD)	77.90%	74.87%	74.84%
ECP (RPL)	72.02%	69.10%	70.40%

Table 5
Forecasting reported as hit rates of actual choices

5. Discussion and conclusion

For junior doctors, the decision of where to practice is important and it will in real life rely on a wealth of information. In consequence, in a stated choice context, they will be surrounded by much uncertainty. Incompleteness of choice scenarios is pervasive in stated choice experiments and generates uncertainty that is expected to be at least in some part resolved at the moment of actual choice. Such resolvable uncertainty is allowed to manifest itself in ECP. In our case, 85% of recorded choices showed some degree of incomplete scenarios. Such information can be used to expand policy options leading to better health care provision, for example by tackling those job factors wrapped in uncertainty.

In terms of modelling, we find that the predictive validity of the LAD model estimated on ECP data is not far from that of the RPL model estimated on conventional DCE data. Our WTP estimates show significant differences between the two preference elicitation methods, as also found by Shoyama et al. (2013). Given that in our data the randomization of respondents across elicitation treatments was successful (except across universities), the observed differences are likely to be due to either the allowance of resolvable uncertainty being explicitly accounted for in the ECP data, or the different econometric model specifications, or the fact that underlying choice behavior changes with the different types of choice tasks posed. The mixed set of results obtained by these initial studies suggests there is some merit in further research based on more in-depth comparison of preference elicitation methods, possibly enriched by advances in psychology. This would help identify the operational and systematic sources of difference and to confirm possible empirical regularities. We report results with remarkable similarities in forecasting performance, despite the LAD model being both a less econometrically demanding approach in terms of estimation than RPL and reliant on less restrictive distributional assumptions on the data generating process. Apart from the standard assumptions for logit probabilities, these are essentially reduced to only symmetry around the median (to overcome the issues of extreme subjective probabilities of 0 and 1 and probability rounding), along with the commonly invoked assumption of zero mean of the idiosyncratic terms in random utility coefficients. No parametric distributional assumption on taste is necessary (e.g. specific shape or size of variance). One could, however, argue that the LAD model should do an even better job in forecasting, since it takes resolvable uncertainty into account, but the reader has to bear in mind that the validity of the LAD model is assessed against a preferred alternative framework, where a correct forecast is obtained only if both the elicited choice probability and the observed choice probability are either below or above 50%. However, forcing the data from the ECP elicitation to fit into the frame of the preferred alternative approach might be too restrictive, as shown by the poor performance in fit of the RPL model when applied to ECP data. Furthermore, a clear limitation of the ECP-LAD approach is that the characterization of preference distributions in the ECP context requires long panels of choice repetitions by respondents (Blass et al. 2010). A further limitation is that it is unclear how to formulate a proper LAD model when choice tasks have more than two alternatives. Nevertheless, our results show that the LAD model predicts with similar accuracy to the conventional preferred alternative preference elicitation approach.

The above limitations with the inherent difficulty of some profiles of respondents to understand and formulate subjective probabilities could hamper the wide-scale adoption of the ECP approach in preference survey research. The latter does not appear to be a problem for the medical students in our sample, although we cannot ignore the evidence of widespread misuse of statistical understanding in clinical practice (Eddy 1982, Cahan et al. 2003). The use of ECP elicitation could be expected to be a problem where the underlying population has a high prevalence of uneducated individuals or more generally for all those with low numeracy skills. We note that this has been investigated within the context of developing countries, where the rate of uneducated or less educated individuals is commonly expected to be large. Both Delavande and Kohler (2009) and Attanasio et al. (2005) find that the basic properties of probabilities are well understood by respondents from rural areas of developing countries. Bruine de Bruin and Fischhoff (2017) showed that more useful responses were elicited using the probabilities compared to using simpler terms, and the numerical probabilities revealed both construct and predictive validity. This was also found to be true for adolescents where the mean probability judgement was found to be close to observed rates. Moreover, when providing visual aids, there is evidence that even young children are able to understand probabilities (for a review on this, see Reyna and Brainerd, 1995). Thus, we argue that it would seem unjustified to disregard a-priori the ECP approach on this basis. We used a probability elicitation scale as a visual aid to respondents. This may have contributed to a greater understanding of the probabilities, which is supported by the fact that fewer respondents chose the limit values of 0 and 1 compared to the study by Blass et al. (2010) where an open-ended question was used. However, in general, the distributions of probabilities were similar in the two studies. Nevertheless, more research on the presentation of different ECP approaches and its impact on choices is warranted.

The use of probabilities can be argued to mimic the real market situation less accurately than discrete choices. Hence the realism of the choice tasks may be greater in the DCE framework as people often have to make discrete choices in actual choice contexts rather than choices based on probabilities. Hence, the ECP approach may provide the analyst with richer data but at the cost of realism. The effect of this limitation in terms of realism is also something that should be investigated further in future comparative work, especially focusing on its consequences in predicting real choice.

Overall, and in line with Manski (1999) and Blass et al. (2010), we argue that using the ECP approach, when contrasted to the standard DCE approach, leads to a better understanding of sources of uncertainty on choice, but with the limitations discussed above. The availability of simpler models that while relying on weaker distributional assumptions provide just as valid predictions is conceptually appealing. Nevertheless, we feel that more research on reliability and validity of WTP estimates, and further comparisons of both probability elicitation scales and econometric approaches are needed to better understand the effects of choice probability elicitation and its link to the underlying choice behavior before one method can generally be recommended over the other in specific contexts of application.

References

Arcidiacono, P., Hotz, V.J., Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals, *Journal of Econometrics*, 166(1):3–16.

Attanasio, O. (2009). Expectations and perceptions in developing countries: their measurement and their use. *American Economic Review* – Papers and proceedings 99(2):87-92.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Blass, A., Lach, S., Manski, C. (2010). Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability, *International Economic Review*, 51(2):421-440.

Brown, P., Panattoni, L., Cameron, L., Knox, S., Ashton, T., Tenbensel, T., Windsor, J. (2015). Hospital sector choice and support for public hospital care in New Zealand: Results from a labeled discrete choice survey, *Journal of Health Economics*, 43:118-127

Bruine de Bruin, W., Fischhoff, B., Millstein, S.G., Halpern-Filsher, B.L. (2000). Verbal and numerical expressions of probability: 'It's a fifty-fifty chance' *Organizational Behavior and Human Decision Processes* 81:115–131.

Bruine de Bruin, W., Fischback, P., Stiber, N., Fischhoff, B. (2002). What number is "fifty-fifty"? Redistributing excessive 50% responses in elicited probabilities. *Risk Analysis*, 22(4):713-723.

Bruine de Bruin, W., Fischhoff, B. (2017). Eliciting probabilistic expectations: collaborations between psychologists and economists, PNAS, www.pnas.org/cgi/doi/10.1073/pnas.1615461114

Cahan, A., Gilon D., Manor O., Paltiel O. (2003). Probabilistic reasoning and clinical decision-making: do doctors overestimate diagnostic probabilities? *Quarterly Journal of Medicine,* 96:763-9.

Caputo, V., Scarpa, R., Nayga, R. and Ortega D. L. (2018). Are preferences for food quality attributes really normally distributed? An analysis using flexible mixing distributions. *Journal of Choice Modelling,* 28:10-27.

ChoiceMetrics. (2009). Ngene 1.0 User manual & reference guide. The cutting edge in experimental design.

Delavande, A., Kohler, H.-P. (2009): Subjective expectations in the context of HIV/ADIS in Malawi. *Demographic Research*, 20(31):817-874.

Dominitz, J., Manski, C. (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, 92(439): 855-867.

Eddy, D.M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Sloviv, Tversky A, eds. *Judgement under uncertainty*: Heuristics and Biases. Cambridge UK, Cambridge University Press, 249–67.

Fischhoff, B., Welch, N., Frederick, S. (1999). Construal Processes in Preference Assessment, *Journal of Risk and Uncertainty* 19(1/3):139-164.

Günther, O.H., Kürstein, B., Riedel-Heller, S.G., König, H.H. (2010). 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

Hensher, D. A., Greene, W.H. (2003). The Mixed Logit model: The state of practice, in: *Transportation* 30(2):133–176

Herriges, J. A., Bhattacharjee, S., Kling, C.L. (2011). Capturing Preferences Under Incomplete Scenarios Using Elicited Choice Probabilities. Working Paper No. 11003 March 2011, Iowa State University, Department of Economics, Ames, Iowa, 50011-1070

Hess, S. (2007). Posterior analysis of random taste coefficients in air travel behaviour modelling. *Journal of Air Transport Management*, 13(4):203-212.

Hess, S. (2010). Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool, *Journal of Choice Modelling*, 3(2).

Holte, J.H., Kjaer, T., Abelsen, B., Olsen, J.A. (2015). The impact of pecuniary and non-pecuniary incentives for attracting young doctors to rural general practice. *Social Science & Medicine*, 128:1-9

Koenker, R., Bassett, Jr. G. (1978). Regression Quantiles, *Econometrica*, 46(1):33-50.

Koenker, R., Hallock K.F. (2001). Quantile Regression, *Journal of Economic Perspectives*. 15(4):143–156

Lancaster, K.J. (1966). A New Approach to Consumer Theory, *Journal of Political Economy* 74 (2):132–157.

Lancsar, E., Wildman, J., Donaldson, C., Ryan, M., Baker, R. (2011). Deriving distributional weights for QALYs through discrete choice experiments, *Journal of Health Economics*, 30(2):466-478

Li, J., Scott, A., McGrail, M., Humphreys, J., Witt, J. (2014). Retaining rural doctors: Doctors' preferences for rural medical workforce incentives. *Social Science & Medicine*, 121:56-64

Luce, R.D. (1959). *Individual choice behavior*, Greenwood Press, Westport, Conn.

Manski, C.F. (2017). Collaboration, conflict, and disconnect between psychologists and economists, *PNAS*, 114(13):3286–3288

Manski, C.F. (2004). Measuring Expectations, *Econometrica*, 72: 1329-1376.

Manski, C.F., (1999). Analysis of Choice Expectations in Incomplete Scenarios. *Journal of Risk and Uncertainty*, 19(1-3):49-66

Manski, C.F. (1990). The Use of Intentions Data to Predict Behavior: A Best Case Analysis. *Journal of the American Statistical Association*, 85:934-940.

Manski, C.F., Molinari, F. (2010). Rounding Probabilistic Expectations in Surveys, *Journal of Business and Economic Statistics*, 28(2): 219-231.

Marchand, C., Peckham, S. (2017). Addressing the crisis of GP recruitment and retention: a systematic review. British Journal of General Practice, 67 (657), e227-e237.

McFadden, D. (1974). Conditional Logit Analysis of Quantitative Choice Behavior, in: Zarembka, P. (Ed.), *Frontiers in econometrics*, Academic Press, New York, pp. 105–142.

Meenakshi, J.V., Banerji, A., Manyong, V., Tomlins, K., Mittal, N., Hamukwala, P. (2012). Using a discrete choice experiment to elicit the demand for a nutritious food: Willingness-to-pay for orange maize in rural Zambia, *Journal of Health Economics*, 31(1):62-71

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis* 9(4): 325-346.

Pedersen, L.B., Gyrd-Hansen, D. (2014). Preference for practice. A Danish study on young doctors choice of general practice using a discrete choice experiment. *European Journal of Health Economics*, 15(6):611-621

Reyna, V., Brainerd, C. (1995). Fuzzy-trace theory: an interim synthesis. *Learning and Individual Differences*, 7:1-75.

Salampessy, B.H. Veldwijk, J., Schuit, A.J., van den Brekel-Dijkstra K., Neslo, R.E.J., de Wit, G.A. Lambooij, M.S. (2015). The Predictive Value of Discrete Choice Experiments in Public Health: An Exploratory Application. *The Patient*, 8: 521–529.

Scanlan G.M., Cleland J., Johnston P., et al. (2018). What factors are critical to attracting NHS foundation doctors into specialty or core training? A discrete choice experiment. *BMJ Open* 2018;8:e019911. doi:10.1136/ bmjopen-2017-019911

Shoyama, K., Managi, S., Yamagata, Y. (2013). Public preferences for biodiversity conservation and climate-change mitigation: A choice experiment using ecosystem services indicators. *Journal of Land Use Policy*, 34:282–293.

Sivey, P., Scott, A., Witt, J., Joyce, C., Humphreys, J. (2012). Junior Doctors' Preferences for Specialty Choice. *Journal of Health Economics*, 31:813-823

Swait, J., Louviere, J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research,* 30:305-314.

Train, K.E. (2009). *Discrete choice methods with simulation*, 2nd Ed., Cambridge University Press, New York.

Wiswall, M., Zafar, B. (2015): Determinants of College Major Choice: Identification using an Information Experiment, *Review of Economic Studies*, 82(2):791-824

Appendix

Choice set 1 of 12	Practice A	Practice B
Population	10,000 - 20,000	5,000 - 10,000
Number of physicians in practice	2	2
Degree of control over working hours	High	Low
Distance to family relations (leisure, school, childcare)	Requires car / public transport	Requires car / public transport
Job security for partner in the community	High	Low
Distance to nearest family	Long drive	Biking distance
Yearly bonus	50,000 DKK	300,000 DKK

Which practice do You prefer?

- Practice A
- Practice B
- I prefer both practices equally

Fig. A1. Example of a discrete choice question

Choice set 1 of 12	Practice A	Practice B
Population	10,000 - 20,000	5,000 - 10,000
Number of physicians in practice	2	2
Degree of control over working hours	High	Low
Distance to family relations (leisure, school, childcare)	Requires car / public transport	Requires car / public transport
Job security for partner in the community	High	Low
Distance to nearest family	Long drive	Biking distance
Yearly bonus	50,000 DKK	300,000 DKK

What is the probability that You choose practice A or B?

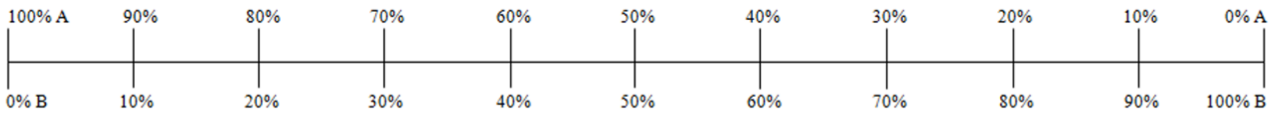

Fig. A2. Example of a choice probability question

Attributes	Levels	Variable labels
Population in the location	Below 2000 inhabitants	Pop2
	2000-5000 inhabitants	Pop25
	5000-10000 inhabitants	Pop510
	10000-20000 inhabitants	Base
Number of GPs in the practice	1 GP (you)	Base
	2 GPs	GP2
	3-4 GPs	GP34
Control over working hours	Low degree	Base
	High degree	Control
Distance to leisure activities, school and day care	Cycling distance	Schoolbike
	Requires car / public transport	Base
Job security for partner in local area	Low	Base
	High	Jobhigh
Distance to closest family	Cycling distance	Familybike
	Short car ride	Familycar
	Long car ride	Base
Yearly bonus	0 DKK	Bonus
	50000 DKK	
	150000 DKK	
	300000 DKK	

Table A1
Overview of attributes and levels

	DCE	ECP	Pearson Chi2 test
Gender			
Men	28%	28%	
Women	72%	72%	0.956
University			
Aarhus University	58%	47%	
Aalborg University	15%	17%	
Copenhagen University	12%	25%	
University of Southern Denmark	15%	12%	0.022
Length of study			
<= 4 years	41%	38%	
> 4 years	59%	62%	0.230
Marital status			
Single	25%	28%	
Married	12%	9%	
Cohabiting	43%	46%	
Have a partner - not cohabiting	20%	16%	
Do not know	1%	1%	0.751
Do you have a state education loan?			
Yes	37%	42%	
No	62%	58%	
Do not wish to disclose	1%	0%	0.284
Do you have, or have you had a study-related job during your education?			
Yes	83%	85%	
No	16%	13%	
Do not wish to disclose	1%	1%	0.639
What specialty do you expect to choose after becoming MD?			
Respondents could choose between 38 different specialties, where general practice was one of them (descriptive stats not displayed here)			0.783
How probable do you consider it to be that you become a general practitioner?			
Very unlikely	10%	11%	
Unlikely	17%	17%	
Neither or	24%	27%	
Likely	35%	28%	
Very likely	15%	17%	0.720
Could you consider taking a job in a rural area without getting economically compensated?			
Yes	52%	49%	
No	28%	27%	
Do not know	20%	24%	0.713

Table A2
Distribution of answers and test for successful randomization

	DCE RPL model Coef. (st.err.) *p*-value	ECP RPL model Coef. (st.err.) *p*-value	ECP LAD model Coef. (st.err.) *p*-value
control	389489	268971	156065
	(49034)	(39387)	(38902)
	<0.001	<0.001	<0.001
familybike	295136	135386	-127116
	(75400)	(49903)	(75615)
	<0.001	<0.001	0.093
familycar	45867	142586	742466
	(127974)	(65809)	(122995)
	0.72	<0.001	<0.001
gp2	-1473800	-36193	240817
	(678029)	(88224)	(116140)
	<0.001	0.68	0.038
gp34	584449	498417	240817
	(121560)	(59852)	(116140)
	<0.001	<0.001	0.038
jobhigh	236521	325424	552286
	(69309)	(53950)	(100581)
	<0.001	<0.001	<0.001
pop2	-73853	-186029	143227
	(107710)	(68603)	(59306)
	<0.001	<0.001	0.016
pop25	-70068	-150497	-1089146
	(93687)	(62016)	(175147)
	<0.001	<0.001	<0.001
pop510	-148780	-94070	-902008
	(34686)	(27613)	(152316)
	<0.001	<0.001	<0.001
schoolbike	361555	227369	-217254
	(49034)	(39387)	(31935)
	<0.001	<0.001	<0.001

Table A3
WTP estimates in DKK