DRAFT – DO NOT QUOTE WITHOUT AUTHORS' PERMISSION

Incorporating "psychological" outcomes into cost-effectiveness analysis

Gerry Richardson[1] and Chris Bojke[2]

1 Centre for Health Economics, University of York

2 Pharmerit UK, York UK

Introduction

Recent policy has targeted self care as a means to improve patient outcomes and reduce costs[1-2]. A recent study[3] demonstrated that an intervention designed to enhance patients' ability to self care was effective at improving the self-efficacy of patients, that is, their confidence in their ability to manage their condition. A separate paper[4] based on the same clinical trial generated QALY gains for these interventions using the EQ5D instrument. The EQ5D instrument measures health related quality of life, across 5 dimensions namely mobility, ability to self care, ability to perform usual activities, level of pain/discomfort and level of anxiety/depression. QALYs generated in this analysis did not explicitly include outcomes such as self-efficacy, which were the primary outcome measure in the clinical study.

Therefore, while QALYs have a commonly expressed value[5], they may not incorporate all the outcomes that are of interest. In contrast, while self-efficacy is undoubtedly an outcome of interest (at least for practitioners and researchers), we have no knowledge of whether self-efficacy is "of value" *per se*, or indeed what that value might be. This leads to problems of interpretation as decision makers cannot assess the relative merits of self-efficacy compared to health related quality of life.

This paper describes a discrete choice experiment (DCE) conducted to examine the relationship between health related quality of life and other "psychological" measures, using self-efficacy as an exemplar. A second component of this paper is an emphasis on identifying whether the econometric results indicate compliance with the underlying theory, Random Utility Theory. Economists often stress the importance of having a theoretical underpinning and though DCE allows scope in assessing whether the theory is consistent with the evidence, this opportunity is largely eschewed in the literature. We address this issue here by focussing on the role of the constant term in the regression model by comparing the theory implied and empirically estimated values and, in addition, note the theoretically inconsistent ways it is often dealt with in the health economics literature. We note that at this stage of the ongoing research we raise more questions than we answer.

Theoretical basis for DCEs

The standard economic model of discrete choice experiments is based on *Random Utility Theory* (RUT) and the *Lancaster's characteristics theory of value*, whereby the indirect utility of consuming a good is a function of the characteristics or attributes of that good and also the individual's characteristics. Thus for any given good the indirect utility function *U(.)* of good *A* with attributes x_A for an individual *i* with characteristics *s* is given by $U_{Ai} = U(x_A; s_i)$. This function is then decomposed in to a potentially observable deterministic component of utility V_{Ai} and a random element ε_A such that $U_{Ai} = V_{Ai} + \varepsilon_{Ai}$.

For simplicity $V(.)$ is often assumed to have a simple additive linear functional form, such that $V_{Aj} = \beta_0 + \beta_1 x_{1A} + ... + \beta_N x_{NA} + \gamma_0 + \gamma_1 s_{1i} + ... + \gamma_M s_{Mi}$. Where x_{nA} is the value of the *n*th characteristic of good *A* and s_{mi} is the value of the *m*th characteristic of individual *i*. β_n and γ_m are the individual invariant (unobserved) parameters related to x_{nA} and s_{mi}, which transform the characteristics of the services and the individual into a utility of that service to that individual. The objective of the DCE is to obtain estimates for the vector of unknown parameters, $(\beta_0, \beta_1, ..., \beta_N)$ (the part-worth utilities). The ratio of any two of these parameters gives the rate of substitution between the variables to which they are associated. Thus if say β_5 is the parameter associated with a cost dimension (money, time or distance) and β_1 is the parameter associated with a higher health state, then β_1 / β_5 represents the relative rate of substitution between the two and therefore may be used to value variables in terms of a cost. ε_{Ai} is assumed to be the standard unobservable error term with the usual properties i.e a random draw from a defined distribution with a mean of 0 and a constant but unknown variance (i.e. homogenous). The error terms are also assumed to be independent of the observed characteristics.

In a DCE where I individuals are asked to make J discrete choices between paired profiles A and B, the paired profiles consists of two goods or services completely described by the vector of variables x_{jA} and x_{jB}. The levels and values of the characteristics of the profiles are varied in each choice according to some experimental design in order to create sufficient variation. Any possible omitted variable or characteristic of each profile is explicitly stated to be constant across all profiles, such that all differences between profile *A* and profile *B* in choice *j* are observed and captured by differences between the vector of variables x_{jA} and x_{jB}.

In each choice the rational respondent will choose the option that yields the highest level of utility to them and so in choice *j* individual *i* will choose B over *A* if $U_{jBi} > U_{jAi}$. From the construction of these indirect utility functions it can be shown that the difference in utility is due to the difference in attributes between option *A* and *B* in choice *j* and the error term.

$$y^*_{ij} = U_{jBi\text{-}jAi} = U_{jBi} - U_{jAi}$$
$$= (\beta_0 + \beta_1 x_{jB} + \ldots + \beta_N x_{NjB} + \gamma_0 + \gamma_1 s_{1i} + \ldots + \gamma_M s_{Mi} + \varepsilon_{jBi}) -$$
$$(\beta_0 + \beta_1 x_{jA} + \ldots + \beta_N x_{NjA} + \gamma_0 + \gamma_1 s_{1i} + \ldots + \gamma_M s_{Mi} + \varepsilon_{jAi})$$
$$= \beta_1 \Delta x_1 + \ldots + \beta_N \Delta x_N + \varepsilon_{jBi} - \varepsilon_{jAi}$$
$$= \beta_1 x_{1j} + \ldots + \beta_N x_{Nj} + \varepsilon_{ji}$$

where $x_{1j} = x_{1jB} - x_{1jA}$ *and* $\varepsilon_{ji} = (\varepsilon_{jBi} - \varepsilon_{jAi})$

As specification individual characteristics appear in equal measure on both sides of the difference equation, they simply difference out and hence the decision to choose A over B or B over A is independent of individuals' observable characteristics. In this *homogenous* model, individuals may value the profiles differently, but the expectation of the difference between the utilities of the two profiles will be the same for all individuals, but will differ in reality subject to the realisation of the error terms.

Econometric Estimation Techniques for DCE

The dependent variable y_{ij}^* is unbounded and continuous, it may be positive indicating a preference for *B* over *A*, or negative indicating a preference for A over B. It may also be zero which would indicate indifference between the two choices. However the analyst does not observe this latent variable but instead observes the outcome, y_{ij} - a binary outcome equal to 1 if option B is chosen and 0 if A is chosen. Thus;

$$y_{ij} = \begin{cases} 1 \text{ if } y_{ij}^* > 0 \\ 0 \text{ otherwise} \end{cases}$$

Taking the previous equation and substituting in for y_{ij}^*, then $y_{ij} = 1 \text{ if } \sum \beta X + \varepsilon > 0$ i.e. $y_{ij} = 1 \text{ if } \sum \beta X > -\varepsilon$. And so the probability that option B is chosen is given by:

$$prob(y_{ij} = 1) = prob(\sum \beta X > -\varepsilon).$$

In order to estimate such a model some assumption is required about the distribution ε and different assumptions about this distribution define the logit, probit and complementary log log models. For example if we assume the error term is iid ~ N(0,1), a draw from the standardised normal distribution, then the probit model is assumed. If the error term were thought to come from standard logistic distribution, then the logit model is assumed. Given an assumption on the behaviour of the error terms, a set of values *X* and a set of 1 or 0 outcomes, the β 's may be estimated.

For example, suppose we had 1000 observations, 975 of them where one (i.e. 97.5%) and 25 where zero and suppose we decided to fit a probit model with just a constant term β_0. Then given our results the proportion of ones and zeros, given the assumption of the error term behaviour and given just a constant term, the model would produce an estimate of $\hat{\beta}_0 = 1.96$. Similarly suppose our observations were split evenly, 500 ones and 500 zeros, then the estimated constant term would be $\hat{\beta}_0 = 0$.

Returning then to the context of the DCE, suppose we give an individual a choice between two identical options A and B, then we would theoretically expect the individual to be indifferent between them i.e. a 50/50 chance of choosing B over A or vice versa. Thus we would clearly expect $\hat{\beta}_0 = 0$ (or at least not significantly different from zero). Any value other than zero, indicates a systematic preference for A over B ($\hat{\beta}_0 < 0$) or B over A ($\hat{\beta}_0 > 0$) all other things being equal. In practice, since on a paper questionnaire, B appears on the right hand side and A on the left, it is plausible that a predominantly right handed population may well systematically choose B over A and give a positive non-zero constant term. However, remember that the $\beta's$ may be compared to each other to establish relative values and trade-offs between dimensions. Thus a relatively large constant term implies respondents are willing to trade substantive benefit for not having to exert the extra effort to tick a left sided box. In other words you would expect there to be a lexicographic preference for all dimensions over the constant term, evidence to the contrary suggests there is something suspect about the underlying premise of the exercise.

The role of the constant term raises its head again when dealing with the fact we observe many observations/choices from individuals. This causes issues with the assumption of independent error terms and the standard response to this in health economics appears to be implementation of a random effects model. However, random effects models account for correlation by allocating each individual their own choice-invariant constant term. Thus random effect models simply allow for heterogeneous preferences for choosing B over A all other things being equal. Thus it is a rather curious solution to a problem as it is inconsistent with the underlying theory and hence it is arguably more of a specification test than a genuine solution.

To summarize the above, it is argued that: An empirically significant constant term is inconsistent with the underlying theory and is a cause for concern and could be evidence of model misspecification and thus, far from being a solution to correlated error terms, significant random effects simply imply a heterogeneous rejection of the underlying model.

We illustrate these arguments with our empirical models following the methods section.

Methods

The authors conducted a Discrete Choice Experiment, a questionnaire based stated choice method, to explore the attributes (or characteristics) that are most valued by patients.

Defining attributes

Qualitative interviews and focus groups carried out alongside a Randomised Controlled Trial of the Expert Patient Programme (EPP) and reviews of the published literature were used to identify the attributes that patients valued. Based on these sources, four attributes were selected for inclusion into the study; health related quality of life (HRQoL), access to General Practitioners, level of isolation and level of self-efficacy (patients' confidence in their ability to manage their condition). Each of the attributes could have three levels. The attributes and levels identified from this process are presented in Table 1.

HRQoL was measured in terms of the EQ5D states. States were selected to ensure a clear ranking between them. Individuals in the DCE were "chronically ill", so including perfect health was considered inappropriate. Three levels were chosen for this attribute to maintain the statistical efficiency of the experiment.

Questionnaire methods.

Two questionnaires were developed to test for both differential response rates and to allow for interactions between attributes. The longer questionnaire contained 28 questions and allowed for interactions on 3 of the 4 attributes, but posed a considerably larger cognitive burden on responders. The shorter questionnaire consisted of 10 questions, but only considered main effects (i.e. it did not allow for estimation of interactions between attributes).

Hypothetical choice sets were created using a fractional factorial design with foldover. Each choice set required the respondent to select one of the unlabelled options A or B (i.e. it was a "forced choice"). The design of both questionnaires was orthogonal, that is main effects could be estimated without correlation with other main effects or interactions[6,7]. An example of the questions facing patients is presented in Table 2.

Consistency

One question in each questionnaire (short and long) was replicated to check for consistency in responses. Thus the short questionnaire contained 9 different questions, while the long questionnaire contained 27 different questions.

Main study

Postal questionnaires were posted to a convenient sample of 511 chronically ill people (255 short questionnaires, 256 long questionnaires). Patients' were randomly allocated to receive either a short or a long questionnaire and a freephone number was provided for any questions patients might have had. Patients who did not respond after two weeks received a written reminder.

Main study sample

Patients who were involved in the Randomised Controlled Trial of the Expert Patients' Programme and had not indicated that they would prefer not to receive any more questionnaires, were included in the study. Thus the study sample was patients with a (self-defined) chronic condition. There were no exclusion criteria and patients from all 28 Strategic Health Authorities were included.

Model estimation

The DCE was analysed by treating each choice between pairs as a single observation. Participants therefore provided either 10 observations if they completed the short questionnaire or 28 observations if they completed the longer questionnaire. The participants' response to each question (i.e. A or B) was included in the model as the binary dependent variable. The independent variables were the differences in the levels of the variables. "Correct" standard errors, allowing for correlation between observations from the same individual were included by clustering on patient identification number.

In the first instance, a standard probit model with constant was used. Subsequently, a random effects probit model with constant was employed with clustering to allow for multiple responses from the same participant. The inclusion of constant terms is a violation of the theoretical basis of the model (see above), but can be used as a notional misspecification test.

The continued identification of a significant constant term lead to other models being considered. There is little guidance in the health economics literature on the procedures for dealing with a significant constant; the problem seems to be largely ignored.

Results

511 patients who had participated in the RCT of the EPP were sent a postal questionnaire. Of these 367 (71.8%) completed and returned one of the questionnaires. Responders were, on average, slightly younger, more likely to own there own home and be in paid employment compared to non-responders. The characteristics of patients who responded compared with non-responders to the questionnaire are presented in table 3. Response rate among those who were sent the shorter questionnaire was higher than for the longer questionnaire (73.7% vs 69.9%), though not substantially so. The characteristics of those returning both types of questionnaire are also presented in Table 3.

Patients' whose responses were inconsistent (in that they gave different answers to the repeated question) were excluded from the primary analysis (n=98, 26.7%). Surprisingly, there were more inconsistent responders to the short questionnaire (n=64, 34.0%), than the long questionnaires (n=34, 19.2%). The remaining 269 responses were considered in the primary analysis.

Variables are dummy coded, with the "best" level of each attribute the omitted category. "Best" in this instance refers to *a priori* expectations that higher levels of health status and confidence would be preferred to lower levels, more frequent visits from friends/relatives would be preferred to fewer and that quicker GP access would be preferred to slower.

<u>Standard probit model with constant term</u>

Results from this analysis are presented in Table 4. The coefficients of each attribute are based on the short and long data combined and represent the impact of a unit increase of each attribute on the probability of getting one outcome (with all the other variables at their mean), and thus the relative importance (marginal rates of substitution) of each attribute can be estimated by dividing one coefficient by another.

The coefficients reflect the disutility associated with moving from one state to another and are intuitively appealing in that the coefficients move in the anticipated direction. The only exception is the movement from isolation level one ("friends and relatives visit daily") to isolation level 2 ("friends and relatives visit every few days"), where the latter is preferred. This movement is not statistically significant and is also plausible in that many individuals with chronic illness would find daily visits burdensome.

As expected, the movement from level one to level three is also greater than movement from level one to two for all attributes.

The coefficients around health status, self-efficacy (confidence) and, to a lesser extent isolation, are of similar magnitude. In particular, the results indicate that patients value a movement from health state one (moderate pain, but no problems on other dimensions) to health state three (moderate pain, moderate anxiety/depression, some problems with self care and some problems with mobility, no problems with usual activities) as approximately the same as a movement from confidence level one ("totally confident in ability to manage condition") to confidence level three ("not at all confident in ability to manage condition") (as the ratio of confidence level 3 to health state 3 is 1.041) .

A difference in "utility" between health states one and three of 0.2507 can be generated from the EQ5D tariff[8]. Thus the movement from not at all confident to totally confident would equate to a gain of 0.25 QALYs if maintained for one year (and assuming that we can ascribe utility values from patient generated responses rather than those of the general public).

Separate analyses for short and long questionnaires were consistent with the combined analysis.

Notice that the constant term included in this model is significant. This constant has no substantive meaning in an unlabelled experiment. In essence, this result implies that responders prefer A to B, even accounting for differences in the scenarios presented to them. Indeed, the size of the coefficient around the constant leads to a more worrying conclusion. If scenarios A and B are identical except that B has health state at level 1 (the "best" health state) and scenario A has health state at level 2 (the moderate health state), A would still be chosen as the coefficient around the constant is greater than that of hl2 (the movement from health 1 to health 2).

Random effects probit

Table 5 shows the results of the random effects probit with constant model. The results are very similar to those from the standard probit model in terms of the magnitude and direction of the coefficients for attributes. The similarity is not surprising as the random effects are attached to individuals and picking up omitted individual characteristics. However, as questionnaires were randomly allocated, there should be no correlation between individuals' characteristics and characteristics appearing in the regression model; therefore omitted variable bias should not occur.

Again however, notice that the constant term is significant.

The health 3 to health 2 ratio in this model was 5.728, while the confidence 3 to health 3 ratio was 1.041, indicating the individuals considered a drop in confidence (from level 1 to 3) as marginally worse that a drop in health state (from 1 to 3), which again, is similar to the results of the standard probit model

Note also that the small values for rho and sigma_u indicate the rejection of the random effects model. This is consistent with the theory stated above.

Generalised Linear and Latent Mixed Models (GLLAMM)

As stated above, we would not expect a constant term to be significantly different from zero. However, in this instance, and in previous evaluations in the health economics literature, the constant term is significantly different from zero. In a labelled experiment, for example where scenario A is always based in a hospital setting, while B is in a home setting, this constant can be argued to reflect the preference for the “label” (in this example a preference for home over hospital or vice versa, other things equal). In an unlabelled experiment, the constant has no meaning and should be zero.

The results of a preliminary GLLAMM model are presented in Table 6.

A basic exploratory latent class model was estimated allowing for heterogeneous preferences across the population. The idea behind the latent class model is that there may be systematic groups who have a systematic set of different preferences. The objective of the latent class regression model is to not only estimate these preferences but also to estimate the proportion of these classes within the population and assign a probability of membership for each individual.

In this example, individuals are in 2 classes. The majority of individuals have a higher probability of being in class one (78%) rather than class 2. It is noticeable that those in class one exhibits the same direction of values of attributes as the previous models (good health is preferred to bad and so on), though the size of these coefficients is larger than previously estimated. For this class of responder, the health 3 to health 2 ratio in this model was 3.285, while the confidence 3 to health 3 ratio was 0.793, indicating that the individuals in this class considered a drop in health state (from level 1 to 3) as considerably worse than a drop in confidence (from 1 to 3).

However, there are a sizable minority (approximately 22%) who have counter-intuitive values particularly for health state. This latent class appear to value decrements in health states whilst still valuing GP access (the other 2 attributes appear much less important to this group).

Health state 3, which is the ‘worst’ health state is preferred to health states 1 and 2. Although this raises some doubts about whether the responders who may be classified as class 2 responders have understood the exercise, it is this class that does not have a significant constant term. This raises the question of what to do with the preferences of this second class – do they genuinely represent odd preferences or are they evidence that a subsection of responders has misunderstood the questionnaire?

In the event that it is the second reason, the effects of miss-responders in the homogeneous model are to weight the preferences of the population towards zero. Note how the ratio of moving down health states and confidence in the ability to manage condition increase relative to GP access (the level which apparently unaffected by the latent class specification).

It should be noted that even though a significant term in a labelled experiment can be explained away by preferences for one or other label, this is not necessarily the only explanation. Heterogeneity of preferences could also explain part of this effect.

<u>Random effects clustered on question number</u>

A further possibility that we wished to explore was whether individuals adopted simple heuristics when options where closely matched. The notion being that if difficult trade-offs were required the respondent may adopt a simple heuristic of choosing right over left. This could be potentially picked up by running a probit model with random effects for each individual question. This will give each of the 38 questions their own specific constant term, a preference for B over A, after allowing for the differences between B and A. The results of this analysis are presented in Table 7.

The ratio of health level 2 to health level 3 is 5.425 and confidence level 3 to health state 3 is 1.044, indicating that a loss of confidence to manage their own condition is worse than a reduction in health state.

Unlike the random effects model which clustered results on individuals, the random effects model clustered on question number indicates that the random effects do indeed explain some of the variation (only a small amount) and that the standard deviation of the question specific constant terms differs from zero. Thus after allowing for what the options contain, some questions are more likely than others to have systematic preferences for B or A.

Given the preferences estimated, the model shows that questions 7 (59% chose B), 1 (57% chose B) and 21 (79% chose B) of the large questionnaire demonstrate an above average and increasing tendency for B to be chosen given the respective levels in A and B. Questions 13 (12% chose B) and 26 (81% chose B) of the large questionnaire and question 7 (76% chose B) of the short questionnaire have an increasing tendency to choose A despite the relative variables in both.

The ‘middle’ questions, i.e. those questions with average random effects are questions 4 (27%) and 6 (64%) of the short questionnaire and question 9 (13%) of the long questionnaire.

Although in this instance, the random effects model appears empirically justified, the effect on the probabilities is relatively small.

Discussion

A response rate to a postal questionnaire of over 70% indicated that patients were willing to complete the DCE. The length, and therefore the cognitive burden of the questionnaire, had little impact on the response rate. Whichever model was chosen, the responses to both questionnaires were consistent with *a priori* expectations. However, the existence of a significant constant term is concerning and the issue of latent classes may require more analysis. The use of random effects probit models may not be the solution to the problem of a significant constant term. One implication of this analysis is that for unlabeled experiments, more work is required to explain the existence of a significant constant. However, a wider implication is that even for labelled experiments, the existence of a significant constant term is worrying. While this has previously been treated as simply a preference for the label (e.g. home over hospital), it is possible that this is not the only explanation.

Nevertheless, the results of this DCE indicate that self-efficacy is an important outcome measure, and patients were willing to trade decrements in HRQoL for improvements in self-efficacy, and that this rate of trade-off was broadly similar across all models. This has implications for the interpretation of economic evaluation in health care. Typically, economic evaluations employ the QALY as a measure of HRQoL. QALYs generated from the EQ5D instrument do not include self-efficacy explicitly and thus QALYs generated in this manner may omit the importance of self-efficacy. Including self-efficacy in the CEA would increase the probability of interventions that improve self-efficacy being cost-effective.

However, there are a number of caveats. Firstly, it is not clear whether self-efficacy is a health outcome; if it is not then should it be included in the objective function of a decision maker working within a budget constrained health system? Should these outcomes be considered? Decision makers need to be clear *a priori* about the maximand and not use vague statements about other factors that may be taken into account.

Secondly, self-efficacy may simply be a measure of process in that patients want self-efficacy as they perceive that this will improve their long term health status. The study surveyed people with chronic conditions and it is likely that these valuations of self-efficacy will be different from those of the general public. Thus, a limitation of this study is that these patients' values are likely to overstate the importance of self-efficacy, and thus make interventions that improve self-efficacy appear more cost-effective.

Thirdly, it is conceivable that changes in self-efficacy (or part of those changes) are already incorporated into the QALY, through one or more of the dimensions of the chosen instrument.

Though self-efficacy has been identified as important to patients, it is likely that in different patient groups other outcomes would be valued and traded for HRQoL. These "important" outcomes should be identified before commencing the study and appropriate techniques should be used to ascertain the rate of substitution between these outcomes and HRQoL.

Finally, it is acknowledged that the use of DCEs may not force responders to focus on the real opportunity cost sacrifice to health by presenting a direct trade-off between health outcome and self-efficacy. This is likely to result in a greater chance of over-stating the relative value of the self-efficacy. Cookson[9] demonstrated that individuals expressed larger relativities when trade-offs were expressed in monetary terms rather than lives saved – similarly in this instance trading–off self-efficacy may be more palatable and therefore result in higher valuations than if individuals were asked to reduce (for example) life expectancy.

Attribute	Levels
Health Related Quality of Life	1 No problems with mobility, usual activities, self care or anxiety/depression. **Moderate pain/discomfort** 2 No problems with usual activities, self care or anxiety/depression. **Some problems with mobility and moderate pain/discomfort** 3 No problems with usual activities. **Some problems with mobility and self care. Also moderate pain/discomfort and moderate anxiety/depression**
Level of confidence	1 Totally confident in ability to manage condition 2 Moderately confident in ability to manage condition 3 Not at all confident in ability to manage condition
Access to General Practitioner	1 GP appointment tomorrow 2 GP appointment in one week 3 GP appointment in 3 weeks
Level of isolation	1 See friends/relatives daily 2 See friends/relatives every few days 3 See friends/relatives rarely

Table 2:

Imagine that you can have either option A or option B, which would you choose? If you would choose the option where you have **moderate** pain or discomfort, are **not** confident that you can manage your condition but you can have a GP appointment **tomorrow** and you see friends or relatives **daily** (i.e. everything in column A) then choose option A.

However, if you would prefer the option where you have **moderate** pain or discomfort as well as having **some** problems with walking, but you are **totally** confident that you can manage your condition with a GP appointment in **3 days time**, but you **rarely** see friends or relatives (i.e. everything in column B), then choose option B.

A		***B***
• You have no problems walking about no problems with self care no problems with usual activities **moderate** pain or discomfort no anxiety or depression • You are **not** confident you can manage your condition • You can have a GP appointment **tomorrow** • You see your friends or relatives **daily**	**OR**	• You have **some** problems walking about no problems with self care no problems with usual activities **moderate** pain or discomfort no anxiety or depression • You are **totally** confident you can manage your condition • You can have a GP appointment in **3 days'** time • You **rarely** see friends or relatives

Please tick one box:

Choice A ☐ **Choice B** ☐

Table 3. Characteristics of non-responders and responders by questionnaire type.

	Non-responders (n=144)	*Responders (n=367) (average)*	*Responders (n=188) (short quest)*	*Responders (n=179) (long quest)*
Characteristic				
Age	52.5	57.5	56.9	58.0
% female	72.9	68.7	73.9	63.1
Accommodation status:				
Owner occupied	57.6	74.4	72.9	76.0
Rented from LA	31.9	18.0	18.1	17.9
Rented privately	8.3	6.0	6.4	5.6
Other	2.1	1.6	2.7	0.6
Condition:				
Musculoskeletal	31.9	37.6	39.9	35.2
Endocrine	13.2	11.2	10.1	12.3
Circulatory	8.3	6.5	4.8	8.4
ME	6.3	7.4	6.9	7.8
Other	40.3	37.3	38.3	36.3
Employment status:				
Employed	16.0	20.4	18.1	22.9
Retired	33.3	39.8	36.7	43.0
Unable to work	36.8	31.6	37.8	25.1
Unemployed	6.3	3.5	4.3	2.8
Other	7.6	4.6	3.2	6.1

Table 4. Probit model with constant term

Attribute	***Coefficient***	***Standard Error***
Health level 2	-0.079	0.033
Health 3	-0.462	0.057
Confidence 2	-0.081	0.024
Confidence 3	-0.468	0.038
GP access 2	-0.072	0.018
GP access 3	-0.254	0.032
Isolation 2	0.001	0.021
Isolation 3	-0.462	0.040
Constant	-0.120	0.017

where
health level 2 is movement from health state one to health state 2
health 3 is movement from health state one to health state 3
confidence 2 is movement from confidence level one to confidence level 2
confidence 3 is movement from confidence level one to confidence level 3
isolation 2 is movement from isolation level one to isolation level 2
isolation 3 is movement from isolation level one to isolation level 3
GP access 2 is movement from GP access level one to GP access level 2
GP access 3 is movement from GP access level one to GP access level 3

Hl3/hl2 = 5.848
cl3/hl3 = 1.041

Table 5. Random effect probit model with constant term

Attribute	***Coefficient***	***Standard Error***
Health level 2	-0.081	0.033
Health 3	-0.464	0.057
Confidence 2	-0.107	0.024
Confidence 3	-0.483	0.038
GP access 2	-0.055	0.018
GP access 3	-0.284	0.032
Isolation 2	0.024	0.021
Isolation 3	-0.426	0.040
Constant	-0.148	0.017

Hl3/hl2 = 5.848
cl3/hl3 = 1.041

Table 6. GLLAMM model

Variable	*Latent Class 1*		*Latent Class 2*	
	Coefficient	Std-errors	Coefficient	Std-errors
Health level 2	-0.288	0.039	0.511	0.071
Health level 3	-0.946	0.058	0.651	0.094
Confidence level 2	-0.243	0.038	0.160	0.063
Confidence level 3	-0.750	0.049	-0.022	0.067
GP Access level 2	-0.017	0.035	-0.174	0.058
GP Access level 3	-0.240	0.044	-0.440	0.072
Isolation level 2	0.015	0.036	0.119	0.064
Isolation level 3	-0.649	0.040	-0.063	0.098
Constant	-0.219	0.028	-0.025	0.045
P(class)	0.784		0.216	

Hl3/hl2 (class 1)= 3.285
cl3/hl3 (class 1)= 0.793
hl3/hl2 (class 2)= 1.274
cl3/hl3 (class 2)= -0.034

Table 7 Random effects on question number

Variable	*Coefficient*	*Std-errors*
Health level 2	-0.087	0.046
Health level 3	-0.472	0.047
Confidence level 2	-0.109	0.046
Confidence level 3	-0.493	0.046
GP Access level 2	-0.052	0.044
GP Access level 3	-0.294	0.049
Isolation level 2	0.029	0.047
Isolation level 3	-0.414	0.046
Constant	-0.154	0.033
Sigma_u	0.161	0.031
Rho	0.025	0.009

hl2/hl3 = 5.425
cl3/hl3 = 1.044

References

1. Department of Health. *Self care - a real choice: self care support - a practical option.* London. Department of Health, 2005.

2. Department of Health. *The Expert Patient: a new approach to chronic disease management for the 21st Century.* HMSO. Department of Health, 2001.

3. Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Richardson G, Gardner C, Gately C and Rogers A. The effectiveness and cost effectiveness of a national lay led self care support programme for patients with long-term conditions: a pragmatic randomised controlled trial

4. Richardson G, Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Gardner C, Gately C and Rogers A. Cost-Effectiveness Analysis of a national lay led self care support programme for patients with long-term conditions.

5. National institute for Health and Clinical Excellence. Guide to the Methods of Technology Appraisal. London: NICE, 2004.

6. Kocur, G., T. Adler, et al. (1982). "Guide to forecasting travel demand with direct utility assessment. US Dept of Transportation, Urban Mass Transportation Administration, Report UMTA-NH-11-0001-82-1."

7. Hahn, G. and S. Shapiro (1966). "A Catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. Report 66-C-165, General Electric REsearch and Development Centre.

8. Dolan P, Gudex C, Kind P, Williams A. A social tariff for Euroqol: Results from a UK general population survey. York Centre for Health Economics Discussion Paper 1995 Sep; 138

9. Cookson R**.** Incorporating psycho-social considerations into health valuations: an experimental study. *Journal of Health Economics* 2000;19:369-401