

# Assessing medicare beneficiaries' strength-of-preference scores for health care options: how engaging does the elicitation technique need to be?

Trafford Crump PhD\* and Hilary A. Llewellyn-Thomas PhD†

\* Candidate and †Professor of Community and Family Medicine, The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School, Hanover, NH, USA

## Abstract

### Correspondence

Trafford Crump  
The Dartmouth Institute for Health  
Policy and Clinical Practice  
35 Centerra Research Park  
Suite 200, Room 210  
Lebanon  
NH 03766  
USA  
E-mail: tcrump@chspr.ubc.ca

### Accepted for publication

15 July 2010

**Keywords:** in-person interview, Medicare population, preference elicitation methods, preference sensitive care, public engagement, strength-of-preferences

**Objective** The objective was to determine if participants' strength-of-preference scores for elective health care interventions at the end-of-life (EOL) elicited using a non-engaging technique are affected by their prior use of an engaging elicitation technique.

**Design** Medicare beneficiaries were randomly selected from a larger survey sample. During a standardized interview, participants considered four scenarios involving a choice between a relatively less- or more-intense EOL intervention. For each scenario, participants indicated their favoured intervention, then used a 7-point Leaning Scale (LS1) to indicate how strongly they preferred their favoured intervention relative to the alternative. Next, participants engaged in a Threshold Technique (TT), which, depending on the participant's initially favoured intervention, systematically altered a particular attribute of the scenario until the participant switched preferences. Finally, they repeated the LS (LS2) to indicate how strongly they preferred their initially-favoured intervention.

**Results** Two hundred and two participants were interviewed (189–198 were included in this study). The concordance of individual participants' LS1 and LS2 scores was assessed using Kendall tau-b correlation coefficients; scores of 0.74, 0.84, 0.85 and 0.89 for scenarios 1–4, respectively, were observed.

**Conclusion** Kendall tau-b statistics indicate a high concordance between LS scores, implying that the interposing engaging TT exercise had no significant effects on the LS2 strength-of-preference scores. Future investigators attempting to characterize the distributions of strength-of-preference scores for EOL care from a large, diverse community could use non-engaging elicitation methods. The potential limitations of this study require that further investigation be conducted into this methodological issue.

## Background

When a society is attempting to provide publicly-funded health services, it inevitably faces the need to prioritize its scarce health care resources.<sup>1</sup> According to the principles of medical ethics, one could argue that efforts to address this need should be carried out in an explicit manner, actively seeking the involvement of a society's members in decisions about priority-setting and resource allocation.<sup>2,3</sup> This argument may be even more relevant in the area of preference-sensitive care, in which there is insufficient clinical evidence to support one optional intervention over all others, or in which there is no clear consensus that a particular intervention's potential benefits overwhelmingly outweigh its potential harms.<sup>4</sup>

To foster the involvement of its members in prioritization efforts, a society would need to collect empirical data about the public's attitudes towards different preference-sensitive health care options. Moreover, that society would need to collect these data in a manner that allows the identification of population sub-groups who, even though they favour the same option, may differ in terms of the strength-of-preference scores they would ascribe to that option.<sup>5</sup>

From a policy perspective, this may be important in order to avoid prioritizing resources and instituting policies that are inadvertently mismatched with the preference-sensitive services that different sub-groups in the public actually want.<sup>4</sup> From an ethical perspective, it may be particularly important to avoid mismatches between what different sub-groups in the public strongly 'want' and what those sub-groups actually 'get' in the context of preference-sensitive care.<sup>3</sup>

### A measurement challenge

These arguments, in turn, imply that we need to be able to collect strength-of-preference scores for health care options from the public. However, several measurement challenges are inherent in efforts to discriminate between relatively strongly- and weakly-held preferences for health care options. The particular challenge addressed

by this paper is whether the designers of future community-wide surveys could, with a fair degree of confidence, use a relatively 'non-engaging' elicitation technique to collect overall strength-of-preference scores for the particular health care options under consideration.

A relatively non-engaging elicitation technique involves: outlining a particular clinical context to the respondent; presenting her with two (or more) therapeutic options for dealing with that situation; asking her to indicate which is her overall favoured option; and then asking her to provide an overall strength-of-preference score for that particular option on some kind of response scale.<sup>6</sup> In large-scale studies involving multiple interviewers who are eliciting strength-of-preference reports from community-dwelling respondents, non-engaging elicitation techniques would be quicker, easier, and less costly to employ than 'engaging' elicitation techniques (which are described below).

However, non-engaging elicitation techniques do not actively involve respondents in making explicit tradeoffs among the different positive and negative attributes that are associated with the alternative health care options.<sup>7</sup> Without such involvement, a respondent may tend to respond to the choice problem without investing much effort in genuinely considering what would be personally at stake in choosing one health care option over another, if she were actually facing this situation in real time. Accordingly, the strength-of-preference scores obtained using a relatively non-engaging elicitation technique may capture a somewhat superficial or unstable picture that is only partially reflective of a participant's actual strength of preference for a particular favoured health care option.

On the other hand, a relatively more engaging elicitation technique does actively involve respondents in making these explicit tradeoffs among the health care options' positive and negative attributes. One could argue that this active involvement heightens the salience of the preference elicitation task, encourages the respondent to invest the effort required to deliberate about the acceptability or unacceptability of the health care options' attributes, and thereby leads the

respondent to gain clearer individualized insight into her own preference structure.<sup>8</sup> Accordingly, the strength-of-preference scores obtained using an engaging elicitation technique may tap into a respondent's strength of preference for a particular favoured option in a deeper, more stable way. If so, then we would wish to use more engaging elicitation techniques – such as the Analytic Hierarchy Process<sup>9,10</sup> or the Threshold Technique (TT)<sup>11,12</sup> – when we try to identify sub-groups of community-dwelling individuals who differ according to the strength-of-preference scores they ascribe to particular health care options. However, these engaging techniques would be slower, more complex, and costlier to employ in population survey study designs than would the non-engaging techniques described above.

#### A strategy

One way to explore this issue is to see what happens when, after considering a particular clinical decision situation, a respondent is: (i) first asked to indicate her strength-of-preference score for a favoured option on a non-engaging Leaning Scale (LS) (a form of category rating scale); then (ii) is asked to interact with an engaging TT (which leads the respondent to consider that favoured option in greater detail, by systematically and repeatedly altering a key attribute of one or the other option and, with each alteration, asking the respondent if she would continue with the favoured option); and finally (iii) is asked again to indicate her non-engaged LS strength-of-preference score for that initially-favoured option.

If the intervening engaging technique actually provides the respondent with deeper insights into her own preferential attitudes towards the options, then we would expect that the strength-of-preference score she reports on the second non-engaging occasion may differ from the strength-of-preference score she reported on the first non-engaging occasion. Such shifts, if they occur, could imply that slower, more complex, and costlier engaging techniques would need to be used in future, large-scale, community-wide surveys assessing the distributions of public

preferential attitudes towards elective therapeutic options.

#### Study purpose

We had an opportunity to carry out this exploratory work, as a follow-up investigation (see the Methods section, below) involving some of the participants in a large, nationally-representative telephone survey of approximately 4000 Medicare beneficiaries in the US.<sup>13–16</sup>

For several reasons, we elected to use an end-of-life decision-making context. During end-of-life care, patients are often provided with interventions that are, at the core, optional in nature and involve difficult tradeoffs between the length and quality of life; their attitudes towards those interventions – and the factors influencing those attitudes – can vary widely.<sup>17–21</sup> Furthermore, the contemplation of these issues is neither a wholly foreign nor a trivial task. There is some evidence that the anticipatory contemplation of preferences about end-of-life care generates reasonable responses from Medicare beneficiaries<sup>18</sup>; therefore, a non-engaging preference elicitation task should at least be feasible for the participants. Also, decisions at the end-of-life are usually one-time events, and it is unlikely that the participants will have repeatedly faced these decisions in the past. If engaging techniques involving trade-offs actually generate greater degrees of salience, encourage participants to gain deeper insights into their own preference structure, and induce an effect on their responses to a subsequent non-engaging approach, we postulated that the end-of-life context would allow that effect to emerge more readily than if we used a clinical context in which the participant encountered more frequently, or in which the consequences were perceived as being less severe (e.g. visiting a physician's office for the flu).

Therefore, the overall purpose of the follow-up study reported here was to address the following methodological research question: When community-dwelling Medicare beneficiaries are asked to consider a hypothetical end-of-life situation with a limited anticipated survival time, to consider two optional actions – one of

which is relatively 'less-intense', while the other is relatively 'more-intense' – and to report, on a non-engaging Leaning Scale, their strength-of-preference scores for the less-intense option, is this strength-of-preference score influenced by the prior use of an engaging Threshold Technique? We explored this research question using the following four different pairs of relatively 'less-intense'/'more-intense' options (see Table 1):

1. No drugs vs. drugs that offer longer survival time but with impaired quality of life.
2. No drugs vs. drugs that offer improved quality-of-life but with shorter survival time.
3. No respirator vs. respirator that offers a 1-month-longer survival time.
4. No respirator vs. respirator that offers a 1-week-longer survival time.

In the Methods section, below, we provide our rationale for focusing our research question on the less-intense therapeutic option in each scenario (see Data Analysis: Preliminary Steps).

## Methods

### Study participants

A large, nationally-representative telephone survey of approximately 4000 Medicare beneficiaries in the US yielded a sampling frame of 427 potential participants who indicated that they would be willing to be approached for a more intensive in-person, follow-up interview.<sup>13–16</sup> Potential participants were located in one of four geographic regions that were deliberately over-sampled in the national telephone survey for the purposes of this in-person interview. Potential participants were considered eligible if they were Medicare beneficiaries, aged 65 or older on July 1, 2003, English- or Spanish-speaking, not institutionalized, and able to provide informed consent to engage in an in-person interview.

### Setting

Participants met face-to-face with a trained community interviewer for a 60-min interview in their home or other mutually agreeable location.

**Table 1** The four end-of-life Decision Scenarios

---

**Decision Scenario #1: No Drugs (Less-Intense Option) vs. Life-Extending Drugs (More-Intense Option)**

---

Suppose that you had a very serious illness. Imagine that no one knew exactly how long you would live, but your doctors said you almost certainly would live less than 1 year.

To deal with that illness, do you think you would want drugs that might lengthen your life beyond 1 year – for about 30 additional days – but would make you feel worse?

---

**Decision Scenario #2: No Drugs (Less-Intense Option) vs. Quality-of-Life Enhancing Drugs (More-Intense Option)**

---

Suppose that you had a very serious illness. Imagine that no one knew exactly how long you would live, but your doctors said you almost certainly would live less than 1 year.

If that illness got to a point that you were feeling bad all the time, do you think you would want drugs that would make you feel better, but might shorten your life by a month?

---

**Decision Scenario #3: No Respirator (Less-Intense Option) vs. Respirator with 1-Month Life Extension (More-Intense Option)**

---

Suppose a year ago you were diagnosed with a very serious illness. Imagine that your doctors had said you almost certainly would live less than a year.

Suppose the year has passed and the illness has got to the point that you needed a respirator to stay alive. If it would lengthen your life for a month, would you want to be put on a respirator?

---

**Decision Scenario #4: No Respirator (Less-Intense Option) vs. Respirator with 1-Week Life Extension (More-Intense Option)**

---

Suppose a year ago you were diagnosed with a very serious illness. Imagine that your doctors had said you almost certainly would live less than a year.

Suppose the year has passed and the illness has got to the point that you needed a respirator to stay alive. If it would lengthen your life for a week, would you want to be put on a respirator?

---

The community interviewer used a standardized, paper interview guide; each participant also had a paper copy of the interview guide. The community interviewer guided the participant through the interview by reading aloud, at the appropriate points in the interview, the text outlining the clinical context, choice scenarios,

stimulus questions, and response choices. The interview schedule also provided the community interviewer with standardized responses to frequently-asked questions that might be posed by the participants (e.g. when presenting the context, the scenarios, or a particular preference elicitation technique). Participants still could indicate if they did not understand the context, scenarios, or questions being asked, or could refuse to answer any of the questions if they wished. The study protocol, including all interview and data-collecting materials, was approved by the Committee for the Protection of Human Subjects at Dartmouth Medical School.

### Presenting the context and the scenarios

At the appropriate point in the interview, the participant was asked to consider being diagnosed with a serious illness and having a prognosis of living no longer than a year. Then four decision scenarios were presented separately (see Table 1). Each involved a pair-wise presentation of two health care options: a relatively less-intense and a relatively more-intense option (i.e. leading to lower or higher demands on health care resources, respectively).

### Data collection

For each decision scenario, the same data collection steps were followed. After considering the scenario and its two relevant therapeutic options, the participant was asked to select the option she would favour, if she were actually in this situation. Then she indicated how strongly she favoured that option – relative to the alternative – on a LS. Next, the interviewer involved the participant in a TT task specifically designed for this study. Finally, the participant was asked to re-consider the scenario and its two relevant therapeutic options, then again indicate how strongly she favoured that option on a new LS. Details about these steps are provided below, using Decision Scenario 1 as an example.

### *The first leaning scale*

The LS was a horizontally-oriented, 7-point bi-directional ordinal scale.<sup>22,23</sup> One therapeutic option appeared at one end of the scale (e.g. in Decision Scenario 1, a score of seven means ‘I strongly favour ‘drugs that may extend the length of life by 30 days but might make me feel worse’). The other therapeutic option appeared at the opposite end (e.g. in Decision Scenario 1, a score of seven means “I strongly favour ‘no drugs’”). There was a labelled neutral point in the middle.

The participant was asked to indicate how strongly she would prefer her favoured option relative to the other, by checking the appropriate point on this scale. Her response was considered her raw, non-engaging ‘time 1’ LS strength-of-preference score (i.e.  $LS_1$ ). See Fig. 1 for an illustrative example.

### *The threshold technique*

After responding to the first LS, the participant worked with the more engaging TT.<sup>11</sup> She was asked to re-consider the same decision scenario, and again to indicate her favoured option. As before, for the purposes of illustration, suppose she was asked to re-consider Decision Scenario 1.

If, in Decision Scenario 1, the participant initially favoured the ‘no drugs’ option, then the 30-day length of life extension offered by the ‘drugs’ option was hypothetically *increased* by regular increments until the participant indicated that she would switch to favouring the initially-rejected ‘drugs’ option. Thus, the participant who initially weakly favoured the ‘no-drugs’ option would switch when there is a slight increase in the length of life extension offered by the ‘drugs’ option, while a participant who initially strongly favoured the same ‘no-drugs’ option would not switch until the length of life extension offered by the ‘drugs’ option is considerably increased.

On the other hand, if, in Decision Scenario 1, the participant initially favoured the ‘drugs’ option, then the 30-day length of life extension offered by the ‘drugs’ option was hypothetically *reduced* by regular decrements until the participant indicated that she would switch to favouring





**Step 1.** The decision scenario is described to the participant along with the two treatment options:  
*“Suppose that you had a very serious illness. Imagine that no one knew exactly how long you would live, but your doctors said you almost certainly would live less than 1 year.*

*To deal with that illness, do you think you would want drugs that might lengthen your life beyond 1 year – for about 30 additional days – but would make you feel worse?”*

**Step 2.** The participant is asked to indicate which treatment option is initially favored:

<b>A: DRUGS</b> They would make you feel worse, but might lengthen your life about <b>30</b> additional days beyond 1 year.	<b>B: NO DRUGS</b> You would feel better than if you took these drugs, but there’d be <b>0</b> additional days of life beyond 1 year.
--	--

**Step 3.** Depending on the initially-favored option, the Community Interviewer (CI) then either systematically increases the length of life offered by the drugs in 5-day increments (if Option B is initially favored) or decreases the length of life offered by the drugs in 5-day decrements (if Option A is initially favored).

**Step 4.** Given the newly-adjusted options, the CI again asks the participant to indicate a stated favored option. If the participant switches to the alternative option, then the CI “fine tunes” in 1-day increments/decrements around this switch point, to see if it could be more closely estimated.

If the participant still favors her initial option, the CI continues to increase or decrease in 5-day increments/decrements until the participant switches to the alternative option. At this point, the CI “fine tunes” in 1-day increments/decrements around this switch point to see if it could be more closely estimated.

If the participant who initially favored Option A would continue to favor Option A even when the additional lifetime offered by the drugs = 0, then the CI double-checks for comprehension. If the participant fully comprehends the exercise, the CI enters “unwilling to trade” (i.e., unwilling to give up drugs).

If the participant who initially favored Option B would continue to favor Option B even when the length of life offered by the drugs > 60 days, then the CI double-checks for comprehension. If the participant fully comprehends the exercise, the CI enters “unwilling to trade” (i.e. unwilling to accept drugs).

**Figure 2** The Threshold Technique (using Decision Scenario #1 as an example).

First, the overall purpose of this study was to determine if there is a method-sequence effect when participants’ strength-of-preference scores are obtained using a non-engaging LS after having used an engaging TT. The second point is that, in order to address this overall purpose in a consistent manner across all the decision scenarios, it was necessary to convert all participants’ raw bi-directional LS scores so that their converted scores – regardless of their initially-favoured option – all lie on the same common, uni-directional underlying strength-of-preference scale. Finally, these converted LS scores could have been oriented towards either the relatively less-intense or the relatively more-intense option; in our conversion, as noted

earlier and as reflected in the study’s exploratory research question, the focus was on the less-intense option.

Accordingly: (i) those who initially-favoured the more-intense option and indicated a raw score for that option of 7, 6, 5, 4, 3, 2, or 1 on the bidirectional LS were assigned a recalibrated score of 1, 2, 3, 4, 5, 6, or 7, respectively; (ii) those who indicated neutral raw scores of 0 on the bidirectional LS were assigned a recalibrated score of 8; and (iii) those who initially-favoured the less-intense option and indicated a raw score for that option of 1, 2, 3, 4, 5, 6, or 7 on the bidirectional LS were assigned a recalibrated score of 9, 10, 11, 12, 13, 14, or 15, respectively,

Thus, in each scenario, all participants' raw LS scores – regardless of their initially-favoured option – were converted so as to yield relative strength-of-preference scores regarding the less-intense option on a new common underlying scale ranging from a score of 1 (a very low strength-of-preference score for the less-intense option) to a score of 15 (a very high strength-of-preference score for the less-intense option).

#### *Analytic steps*

Those participants who did not understand the clinical context, could not indicate an initially-favoured option, or refused to answer the questions, were excluded from the analysis, and a response rate was calculated for each separate scenario. For each scenario, descriptive statistics were used to summarize: the sub-study participants' characteristics; the frequencies at which the less-intense and more-intense options were initially-favoured; the responses to the TT; and the distributions of the *converted* LS<sub>1</sub> and LS<sub>2</sub> strength-of-preference scores regarding the less-intense option.

Then, for each scenario, the intra-participant agreement across the *converted* LS<sub>1</sub> and LS<sub>2</sub> strength-of-preference scores for the less-intense option was assessed using the Kendall tau-b correlation coefficient (Kendall's tau). This nonparametric test can be used to assess the level of concordance between paired, ordinal-level variables (i.e. the *converted* LSs' scores) for a specific observation (i.e. each study participant). Kendall's tau coefficient values can range from -1.00 to 1.00. A value of -1.00 indicates perfect disagreement between the

variables, while a value of 1.00 indicates perfect agreement between the variables. It can be interpreted as the probability of observing either a concordant ( $0 \leq t \leq 1$ ) or discordant pair ( $-1 \leq t \leq 0$ ).<sup>24</sup>

## Results

### Response rates and participant characteristics

Among the 202 participants, response rates for the four end-of-life decision scenarios ranged from 94% ( $n = 189$ ) to 98% ( $n = 198$ ). The participants' average age was 76 years; the overall sample was 54% female, 92% were white, 87% were English-speaking, 83% had at least a high school education, 62% were married, and 75 and 90% considered their physical and mental health, respectively, to be 'good' to 'excellent'.

### Initially-favoured options

The majority of study participants initially favoured the less-intense option in three of the four scenarios; see Table 2.

### Strength-of-preference scores on the unidirectional leaning scale

Recall that, according to the *converted* unidirectional LS scoring strategy, '1' is a low strength-of-preference score and '15' is a high strength-of-preference score for the less-intense option. In all four decision scenarios, the converted LS<sub>1</sub> and LS<sub>2</sub> scores ranged from 1 and

**Table 2** Initially favour less-intense option, initially favour more-intense option, or indifferent between options: frequencies, by Decision Scenario ( $N = 202$ )

Decision Scenario	No drugs	Indifferent	Drugs
Scenario #1: Life-extending drugs vs. no drugs ( $n = 198$ )	180 (91%)	6 (3%)	12 (6%)
Scenario #2: Quality-of-life enhancing drugs vs. no drugs ( $n = 189$ )	45 (24%)	10 (5%)	134 (71%)
	No Respirator	Indifferent	Respirator
Scenario #3: Respirator with 1-month Life extension vs. no respirator ( $n = 194$ )	167 (86%)	4 (2%)	23 (12%)
Scenario #4: Respirator with 1-week life extension vs. no respirator ( $n = 198$ )	170 (86%)	6 (3%)	22 (11%)



**Table 3** Participants reporting a shift from Time 1 (LS<sub>1</sub>) to Time 2 (LS<sub>2</sub>) in unidirectional Leaning Scale scores for the less-intense option: frequency counts and percentages, by Decision Scenario

Decision Scenario (N)	Total number Reporting any Shift (%)	Number reporting higher scores at Time 2	Number reporting lower scores	Number reporting a single-point shift
Scenario #1 (n = 198)	47 (23.7%)	26 (mean increase = 2.42)	21 (mean decrease = 2.19)	23 of 47
Scenario #2 (n = 189)	40 (21.1%)	20 (mean increase = 2.20)	20 (mean decrease = 2.35)	19 of 40
Scenario #3 (n = 194)	41 (21.1%)	22 (mean increase = 1.86)	19 (mean decrease = 1.84)	25 of 41
Scenario #4 (n = 198)	22 (11.1%)	6 (mean increase = 2.50)	16 (mean decrease = 1.67)	9 of 22

15, respectively. In three of the four decision scenarios (Scenarios #1, #3, and #4), the distributions of these converted LS<sub>1</sub> and LS<sub>2</sub> scores were dominated by high strength-of-preference scores for the less-intense option, with modal scores of 15. However, different distributions were observed for Decision Scenario #2. For this decision scenario, the distributions of the converted LS<sub>1</sub> and LS<sub>2</sub> scores were dominated by low strength-of-preference scores for the less-intense option ('no drugs'), with a modal score of 1.

The frequency counts of the participants who reported different LS scores at Time 1 and Time 2 are provided in Table 3. For Decision Scenarios #1, #2, #3 and #4, across-time shifts in the converted LS<sub>1</sub> and LS<sub>2</sub> scores were observed for 24, 21, 21 and 11% of the participants, respectively. Of those participants in whom shifts were observed, 49, 48, 61 and 41% (for Decision Scenarios #1–#4, respectively) changed their strength-of-preference score by only a single point in either direction.

#### What happened between LS<sub>1</sub> and LS<sub>2</sub>?

This study primarily focused on whether or not the TT (as one example of an intervening, engaging elicitation technique) generated differential effects on the converted LS<sub>2</sub> scores (as an example of a non-engaging elicitation technique). Accordingly, the TT results are not in themselves of primary interest here. However, in Table 4 we summarize these observations, in order to fully report the results obtained at all three preference-elicitation points.

#### Answering the research questions

For each of the four exploratory research questions, the intra-participant agreement across the converted LS<sub>1</sub> and LS<sub>2</sub> scores was assessed using Kendall's tau. Kendall tau-b correlation coefficients of 0.74, 0.84, 0.85 and 0.89 were observed for each of the four scenarios, respectively; for all coefficients,  $P < 0.01$ .

#### Discussion

When health care researchers aim to characterize the distributions of public attitudes towards different elective health care options, they may wish to use methods that indicate how *strongly* different members of the public appear to hold these preferences. Consider, for example, the perspective of health services researchers who want to identify unwarranted variations in preference-sensitive care.<sup>13,25</sup> If health services researchers could clearly identify sub-groups of community-dwelling individuals who differ according to the strength of their preferences for particular health care options, then their attempts to design, test, and implement interventions to reduce unwarranted variations in preference-sensitive care could be more efficiently targeted, in that they could concentrate on areas where the gaps between individuals' health care preferences and the care that they actually receive are widest.<sup>26</sup>

Before such studies can be conducted, however, several methodological issues have to be addressed. This study focused on one of those issues: how engaging does the preference elicitation technique 'need' to be?

**Table 4** Threshold Technique switch points, by Decision Scenario (*N* = 202)

Decision Scenario ( <i>n</i> )	TT procedure	Results
<u>Scenario #1</u> ( <i>n</i> = 198) 'No drugs with anticipated survival of 12 months' vs. 'drugs with 30-day lifetime extension, so anticipated survival is 13 months but with impaired quality of life'	If initially favours 'no drugs', then increase lifetime extension with the 'drugs' option until respondent switches to 'drugs'	For 143 (72%), no amount of lifetime extension would lead to accepting these drugs.
	If initially favours 'drugs', then decrease lifetime extension with the 'drugs' option until respondent switches to 'no drugs'	For 49 (25%), minimal required lifetime extension from these drugs: mean = 208 days; median = 180 days.
<u>Scenario #2</u> ( <i>n</i> = 189) 'No drugs with anticipated survival of 12 months' vs. 'drugs with improved quality of life but with 30-day reduction in lifetime, so anticipated survival is 11 months'	If initially favours 'no drugs', then decrease lifetime reduction with the 'drugs' option until respondent switches to 'drugs'	For 38 (20%), maximal tolerable lifetime reduction from these drugs = 0
	If initially favours 'drugs', then increase lifetime reduction with the 'drugs' option until respondent switches to 'no drugs'	For 141 (75%), maximal tolerable lifetime reduction from these drugs: mean = 100 days; median = 50 days.
<u>Scenario #3</u> ( <i>n</i> = 194) 'No respirator with anticipated survival of 12 months' vs. 'respirator with 30-day lifetime extension, so anticipated survival is 13 months but with impaired quality of life'	If initially favours 'no respirator', then increase lifetime extension with the 'respirator' option until respondent switches to 'respirator'	For 155 (80%), no amount of lifetime extension would lead to accepting this respirator.
	If initially favours 'respirator', then decrease lifetime extension with the 'respirator' option until respondent switches to 'no respirator'	For 32 (16%), minimal required lifetime extension from this respirator: mean = 137 days; median = 75 days
<u>Scenario #4</u> ( <i>n</i> = 198) 'No respirator with anticipated survival of 12 months' vs. 'respirator with 7-day lifetime extension, so anticipated survival is 12 months + 7 days but with impaired quality of life'	If initially favours 'no respirator', then increase lifetime extension with the 'respirator' option until respondent switches to 'respirator'	For 157 (79%), no amount of lifetime extension would lead to accepting this respirator.
	If initially favours 'respirator', then decrease lifetime extension with the 'respirator' option until respondent switches to 'no respirator'	For 35 (18%), minimal required lifetime extension from this respirator: mean = 94 days; median = 30 days.

Our exploratory work used four decision scenarios pertaining to an end-of-life context. Each scenario offered a choice between a relatively less- or more-intense therapeutic option. After the participant indicated an initially-favoured option, she first indicated, on a relatively non-engaging bi-directional LS, how strongly she preferred that initially-favoured option, then she worked with a relatively engaging TT, and, finally, she indicated again, on a new non-engaging bi-directional LS, how strongly she preferred her initially-favoured option. Subsequently, in order carry out the analysis required to address our four research questions, we converted all raw bi-directional LS scores to uni-directional strength-of-preference scores oriented towards the less-intense option. In all four scenarios, we observed a high degree of concordance between participants' non-engaging converted LS<sub>1</sub> and LS<sub>2</sub> strength-of-preference-scores for the less-intense option, in spite of the intervening engagement with a scenario-specific TT.

This observation is noteworthy, given others' observations about the construction of preferences during the process of preference elicitation, under controlled experimental conditions. For example, experiments in psychology have implied that the nature of the preference elicitation task can itself influence a preference report.<sup>8</sup> Furthermore, when an individual is faced with a decision problem that is important, complex, and infrequent – comparable to many health care decision situations – the manner in which that decision problem is presented can generate phenomena such as preference reversals.<sup>27,28</sup> It is postulated that this occurs because the presentation of the decision problem and the preference elicitation task themselves act as values clarification interventions, encouraging participants actively to work through various attributes of a decision problem – for example, to engage in active risk appraisal and evaluation – which, in turn, fosters the construction and report of previously-unformulated preferences.<sup>29,30</sup> However, our observations – albeit in a 'real-world', community-based interview as opposed to a controlled experiment – are not consistent with these postulations.

### Limitations

The high degrees of concordance that we observed may actually be a form of Type II error, generated by four different study design issues.

First, perhaps we sampled participants who, prior to their recruitment into this study, had already formulated their preferences for the care they would like to receive at the end of life. If this were the case, the TT – or any other engaging elicitation technique – would have minimal influence on the subsequent strength-of-preference LS scores, and thus the ability to detect any method-sequence effects would be limited. On the other hand, research into the completion of advanced directives indicates that as few as 4–25% of specific patient groups have provided formal instructions for their care at the end-of-life, suggesting that the prior formulation of such preferences actually may not be widely prevalent amongst the elderly individuals involved in our study.<sup>31</sup> In any case, there is still a need to use other, non-end-of-life, preference-sensitive decision scenarios in future investigations into possible method-sequence effects.

Second, it is possible that the TT, as it was designed for this particular application, actually does not encourage participants actively to engage in the construction and reporting of previously-unformulated preferences, as postulated by Fischhoff.<sup>29,30</sup> If we had, instead, used a different engaging preference elicitation technique – such as the Balance Technique with the LS,<sup>22</sup> or the Analytic Hierarchy Process<sup>10</sup> – we might have observed a notable method-sequence effect.

Third, when we used the TT, it is possible that we worked with a scenario attribute that was not actually very salient to our participants. Recall that the attribute used here – overall survival time – was approached in two different ways; in Scenarios #1, #3 and #4 we worked with lifetime extension (with a loss in quality-of-life), and in Scenario #2 we worked with lifetime reduction (with a gain in quality-of-life). If the attribute of overall survival time was, in fact, immaterial to our study participants, then engagement with this version of the TT would not stimulate shifts

– either one way or the other – in the LS<sub>2</sub> scores. On the other hand, versions of the TT that worked with a more salient scenario attribute – for example, the levels of pain relief offered by an intervention – might generate shifts in the LS<sub>2</sub> scores.<sup>32</sup>

Finally, it is possible that either the act of indicating an overall initially-favoured option or the act of providing a ‘time 1’ LS score can, in itself, create a kind of *preferential* anchoring effect. Strictly speaking, an anchoring effect is generated in a *probabilistic* context.<sup>33</sup> An individual is provided with an objective statement about the likelihood of a ‘reference’ event’s occurrence, and then is asked to provide her own subjective estimates of the probability of other, different, events. Her subjective estimates can be systematically affected by the level of probability appearing in the original likelihood statement. If this anchoring phenomenon is also strongly present in preference-elicitation work, then it is going to be very difficult to devise study designs for investigating method-sequencing effects. There may be ways to offset this kind of artifactual effect. For example, investigators could lengthen the time interval between the preference elicitation tasks, and could avoid revealing to participants the responses they had provided on the first LS. However, the introduction of these kinds of design controls may actually distort the preference-elicitation interview either into an experience in which the respondent assumes that she ‘must’ provide consistent across-technique responses, or into a format that’s not feasible in a large-scale community survey.

## Conclusion

The lack of an apparent method-sequence effect could imply that future community surveys aimed at collecting empirical data about the public’s preferential attitudes towards differing health care options in an end-of-life context may be able to proceed using the more tractable ‘non-engaging’ techniques. However, given the potential limitations to the work reported here, further methodological research is needed to more fully investigate the ways in which

respondent characteristics, study designs, and the elicitation methods themselves interact, before such an implication could be conclusively accepted.

## Acknowledgements

Crump T, Llewellyn-Thomas H. *Eliciting Medicare Beneficiaries’ Strength of Preference Scores for Elective Health Care Interventions at the End-of-Life: Is a ‘Non-engaged’ Approach Good Enough?* Podium presentation at the 5th bi-annual International Shared Decision Making conference. June, 2009, Boston, MA, USA.

## Conflicts of Interest

None.

## Sources of Funding

National Institutes of Health: National Institute of Aging and Foundation for Informed Medical Decision Making.

## References

- 1 Baicker K, Chandra A. Uncomfortable arithmetic – Whom to cover versus what to cover. *New England Journal of Medicine*, 2010; **362**: 95–97.
- 2 Lauridsen S, Norup M, Rossel P. The secret art of managing healthcare expenses: investigating implicit rationing and autonomy in public healthcare systems. *Journal of Medical Ethics*, 2007; **33**: 704–707.
- 3 Owen-Smith A, Coast J, Donovan J. Are patients receiving enough information about healthcare rationing? A qualitative study. *Journal of Medical Ethics*, 2010; **36**: 88–92.
- 4 Wennberg J, O’Connor A, Collins E, Weinstein J. Extending the P4P agenda, part 1: how Medicare can improve patient decision making and reduce unnecessary care. *Health Affairs*, 2007; **26**: 1564–1574.
- 5 Ryan M, Scott D, Reeves C, Bate A, van Teijlingen E, Russell E. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment*, 2001; **5**: 1–186.
- 6 Llewellyn-Thomas H. Values clarification. In: Edwards A, Elwyn G (eds) *Shared Decision-Making in Health Care: Achieving Evidence-Based Patient Choice*, 1st edn. New York, USA: Oxford University Press, 2009: 123.

- 7 Keeney R, Raiffa H. *Decisions with Multiple Objectives Preferences and Value Tradeoffs*. New York, USA: Cambridge University Press, 1993.
- 8 Lichtenstein S, Slovic P. The construction of preferences: an overview. In: Lichtenstein S, Slovic P (eds) *The Construction of Preferences*, 1st edn. New York, USA: Cambridge University Press, 2006: 1–40.
- 9 Dolan J, Isselhardt B Jr, Cappuccio J. The Analytic Hierarchy Process in medical decision making: a tutorial. *Medical Decision Making*, 1989; **9**: 40–50.
- 10 Dolan J. Shared decision-making – Transferring research into practice: The Analytic Hierarchy Process (AHP). *Patient Education and Counseling*, 2008; **73**: 418–425.
- 11 Llewellyn-Thomas H. Threshold Technique. In: Kattan MW (ed.) *Encyclopedia of Medical Decision Making*. Thousand Oaks, CA: Sage Publications, 2009: 1134–1137.
- 12 Llewellyn-Thomas HA. Investigating patients' preferences for different treatment options. *Canadian Journal of Nursing Research*, 1997; **29**: 45–64.
- 13 Anthony D, Herndon B, Gallagher P *et al*. How much do patients' preferences contribute to resource use? *Health Affairs*, 2009; **28**: 864–873.
- 14 Barnato A, Anthony D, Skinner J, Gallagher P, Fisher E. Racial and ethnic differences in preferences for end-of-life treatment. *Journal of General Internal Medicine*, 2009; **24**: 695–701.
- 15 Barnato A, Herndon B, Anthony D *et al*. Are regional variations in end-of-life care intensity explained by patient preferences? *Medical Care*, 2007; **45**: 386–393.
- 16 Fowler F Jr, Gallagher P, Anthony D, Larsen K, Skinner J. Relationship between regional per capita Medicare expenditures and patient perceptions of quality of care. *Journal of the American Medical Association*, 2008; **299**: 2406–2412.
- 17 Rosenfeld K, Wenger N, Kagawa-Singer M. End-of-life decision making: a qualitative study of elderly individuals. *Journal of General Internal Medicine*, 2000; **15**: 620–625.
- 18 Rosenfeld K, Wenger N, Phillips R *et al*. Factors associated with change in resuscitation preference of seriously ill patients. The SUPPORT investigators study to understand prognoses and preferences for outcomes and risks of treatments. *Achieves of Internal Medicine*, 1996; **156**: 1558–1564.
- 19 Fried T, Van Ness P, Byers A, Towle V, O'Leary J, Dubin J. Changes in preferences for life-sustaining treatment among older persons with advanced illness. *Journal of General Internal Medicine*, 2007; **22**: 495–501.
- 20 Fried T, Byers A, Gallo W *et al*. Prospective study of health status preferences and changes in preferences over time in older adults. *Archives of Internal Medicine*, 2006; **166**: 890–893.
- 21 Fried T, Bradley E. What matters to older seriously ill persons making treatment decisions? A qualitative study. *Journal of Palliative Medicine*, 2003; **6**: 237–244.
- 22 O'Connor A, Jacobsen M, Stacey D. An evidence-based approach to managing women's decisional conflict. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 2002; **31**: 570–581.
- 23 Bech M, Gyrd-Hansen D, Kjaer T, Lauridsen J, Sorensen J. Graded pairs comparison – Does strength of preference matter? Analysis of preferences for specialised nurse home visits for pain management. *Health Economics*, 2007; **513**: 513–529.
- 24 Abdi H. The Kendall rank coefficient. In: Salkind N (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA: Sage Publications, 2007: 508–510.
- 25 Wennberg J, Fisher E, Skinner J. Geography and the debate over Medicare reform. *Health Affairs*. Web Exclusive 2002;W96–W114.
- 26 O'Connor A, Llewellyn-Thomas H, Flood A. Modifying unwarranted variations in health care: shared decision making using patient decisions aids. *Health Affairs*. Web Exclusive 2004;var 63–72.
- 27 Slovic P. The construction of preference. *American Psychologist*, 1995; **50**: 364–371.
- 28 Johnson E, Steffel M, Goldstein D. Making better decisions: from measuring to constructing preferences. *Health Psychology*, 2005; **24**: S17–S22.
- 29 Fischhoff B. Values elicitation: is there anything there?. In: Hechter M, Nadel L, Michod R (eds) *The Origins of Values*. New York, USA: Aldine Transaction, 1993: 187–214.
- 30 Fischhoff B. Constructing preferences from liable values. In: Lichtenstein S, Slovic P (eds) *The Construction of Preferences*, 1st edn. New York, USA: Cambridge University Press, 2006: 653–667.
- 31 Perkins H. Controlling death: the false promise of advance directives. *Annals of Internal Medicine*, 2007; **147**: 51–57.
- 32 Kopec J, Richardson C, Llewellyn-Thomas H, Klinkhoff A, Carswell A, Chalmers A. Probabilistic Threshold Technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied. *Journal of Clinical Epidemiology*, 2007; **60**: 929–938.
- 33 Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*, 1974; **185**: 1124–1131.