(Running Head: The impact of labelling on health state values)

Title: 'Naming and Framing': The impact of labelling on health state values for multiple sclerosis

Colin Green (PhD) [1,2] *, Elizabeth Goodwin (PhD) [1], Annie Hawton (PhD) [1,2]

[1] Health Economics Group, University of Exeter, Exeter, UK, EX1 2LU

[2] South West Collaboration for Leadership in Applied Health Research and Care (CLAHRC), University of Exeter Medical School, University of Exeter, Exeter, UK, EX1 2LU

***** *Corresponding Author: Colin Green, University of Exeter, St Luke's Campus, South Cloisters, Exeter, EX1 2LU. Email: c.green@exeter.ac.uk*

Word-count total for the manuscript text = 4,300 (excluding the abstract, acknowledgments, references, and figure legends).

Financial Statement: Financial support for this study was provided in part by the Multiple Sclerosis Society of Great Britain and Northern Ireland. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. C Green and A Hawton acknowledge partial funding from the UK NIHR Collaboration for Leadership in Applied Health Research and Care of the South West Peninsula (PenCLAHRC). The views expressed in this publication are those of the authors and not necessarily those of the Multiple Sclerosis Society, the UK NIHR or the Department of Health.

Abstract

Introduction: Health state valuation is a key input to many economic evaluations that inform resource allocation across competing health care interventions. Empirical evidence has shown that respondents to preference elicitation surveys may value a health state differently if aware of the condition causing it ('labelling effects'). This study investigates the impact of including a multiple sclerosis (MS) label for valuation of MS health states.

Methods: Health state values for MS were elicited using two internet-based surveys in representative samples of the UK population (n=1702; n=1788). In one survey respondents were not informed that health states were caused by MS. The second survey included a condition label for MS. Surveys were identical in all other ways. Health states were described using a MS-specific eight-dimensional classification system (MSIS-8D), and the time trade-off valuation technique was used. Differences between values for labelled and unlabelled states are assessed using descriptive statistics and multivariate regression methods.

Results: Adding a MS condition label had a statistically significant effect on mean health state values, resulting in lower values for labelled MS states versus unlabelled states. Data suggests the MS label had a more significant effect on values for less severe states, and no significant effect on values for the most severe states. The inclusion of the MS label had a differential impact across the dimensions of the MSIS-8D. Across the MSIS-8D, predicted values ranged from 0.079-0.883 for unlabelled states, and 0.066-0.861 for labelled states.

Conclusions: Differences reported in health state values, using labelled and unlabelled states, demonstrate that condition labels affect the results of valuation studies, and can have important implications in decision-analytic modelling and in economic evaluations.

1. Introduction

Economic evaluations in the health care setting commonly employ quality-adjusted life-years (QALYs), within cost-utility analyses, to inform decisions on the reimbursement of health care interventions. The basic QALY construct is that individuals experience different health states over time and that each health state has a value attached to it, with QALYs used to represent the value-weighted (quality-adjusted) accumulated time e.g. life-years[1]. Health state values (HSVs) are on a scale where the value of being dead is zero (0), and the upper end of the scale has a value of one, reflecting the best imaginable health state (or perfect health, or similar description). States worse than dead are possible and have a negative value. HSVs reflect the relative preferences of respondents across different health states, whereby a health state description that is more (or less) desirable than others will have a value that is closer to (or further away from) one (1.00) than the comparison health states.

Although the use of the QALY in the assessment of health benefits in economic evaluations continues to be a major source of contention[1, 2], the QALY is widely used and is the recommended outcome measure in a number of policy settings[3-6]. There are variations in the estimation of health state values and the major areas of uncertainty can be characterised in three broad methodological areas; what to value? (i.e. how to describe health states), how to elicit preferences (values)?, and from whom to elicit preferences (whose values)? Here we are concerned with the first of these questions, when considering whether or not to name the disease or health condition in the health state description being valued.

Empirical evidence suggests that values for the same health states can vary depending upon whether or not the underlying condition is 'labelled', i.e. made known to respondents[7]. While some studies provide evidence that including labels for particular conditions significantly reduces HSVs, others have found no effect, and the results vary between conditions[7-11]. In a recent study[7] the use of a label for cancer health states, compared to otherwise identical unlabelled health states, resulted in significantly lower HSVs for severe health states but did not affect valuations of milder states. There are alternative normative arguments for and against the use of condition labels. For example, the condition itself may have an effect on HRQoL that is unrelated to a generic health state description. Here Brazier and colleagues (2012)[12] have drawn attention to potentially different valuations when people have 'some difficulty in taking a long walk' when it is due to needing to be near a toilet or owing to psychological problems, versus having 'some difficulty taking a long walk' due to physical problems. In this case, condition labels may result in more accurate HSUVs. Alternatively, respondents may be influenced by their own knowledge, experience or preconceptions about the condition, in which case it may be argued that HSUVs could be

distorted due to irrelevant or inaccurate factors. However, there is little empirical evidence to explain such differences in the values ascribed to labelled and unlabelled health states, and consensus has yet to be been reached about the extent to which condition labels influence HSVs.

A number of condition-specific preference based measures (CSPBMs) have been developed[13], for use as alternatives to, or in addition to generic PBMs (e.g. EQ-5D[14], SF-6D[15]). Yet, even when using a CSPBM the current guidance suggests that until more is known about which condition labels affect HSVs, and why discrepancies occur, condition labels should be avoided[12]. Given the apparent variability of labelling effects between conditions, further research has been recommended to investigate the impact of labels in studies valuing states for particular conditions[7, 12].

In this study, the development of a condition-specific PBM for multiple sclerosis (MS) provided an opportunity to investigate the impact of including a condition label on public valuations of health states for MS, a chronic condition with a wide variety of symptoms. As far as we are aware, this study is the first to derive a full tariff of values for labelled and unlabelled versions of the same classification system.

2. Methods

Two surveys were undertaken to elicit preferences for MS health state descriptions. One avoided any mention of MS (the unlabelled version of the survey, Survey 1), while the other explicitly stated that the health states described the impact of MS (the labelled version, Survey 2). In order to ensure that the results of the two surveys are comparable, the same preference elicitation methods were used for both surveys, and the same approach used for descriptive statistics and statistical analyses, consistent with current methodological guidance[16].

The MS health states were described using the Multiple Sclerosis Impact Scale Eight Dimension (MSIS-8D) classification system[17] (Figure 1), which was derived from the MSIS-29 item patient reported outcome measure, a frequently used and well validated measure of health-related quality of life (HRQoL) in MS[18, 19].

< Insert Figure 1 >

The development of the MSIS-8D classification system is described in detail elsewhere[17]. The measure comprises eight dimensions of importance to the HRQoL of people with MS: physical functioning, mobility, social activities, daily activities, fatigue, cognitive function,

emotional well-being and depression. Each dimension is represented by one MSIS-29 item, and each item has four response levels: (1) not at all, (2) a little, (3) moderately and (4) extremely.

The survey methods are described in more detail elsewhere[20]. In summary, a sample of 169 MSIS-8D health states was selected to reflect states that are likely to be experienced by people with MS at different levels of severity. These health states were allocated to five severity groups according to the sum of their levels across dimensions (see Supplementary material). Each respondent valued one set of five MSIS-8D health states, covering a range of condition severity - the survey design comprised 34 sets of five health states - plus the worst health state described by the MSIS-8D (the pits state). Prior to preference elicitation respondents completed the MSIS-8D for their own health, ranked three MSIS-8D health states and completed a practice TTO task. Preferences were elicited using an internet version of the Measurement and Valuation of Health (MVH) variant of the time trade-off (TTO) technique[21], which asks respondents to state whether they would prefer ten years in the target health state or a shorter period of time (x) in perfect health. The length of time spent in perfect health (x) is varied until the respondent is indifferent between the two options, and the HSV is calculated as $x/10$. Cognitive testing and an online pilot study (n=50) were used prior to the full survey to ensure the online survey worked as intended. The MVH protocol for health state valuation, and the elicitation of preferences from a representative sample of the UK population, were applied here to be consistent with recommendations from the UK National Institute for Health and Care Excellence (NICE) on methods for economic evaluations[3].

The online surveys were programmed and hosted by Accent Marketing and Research Ltd (http://www.accent-mr.com/), with participants recruited from an existing internet panel held by Survey Sampling International (SSI), an experienced online panel provider with rigorous quality control procedures. Each survey aimed to obtain data from 1700 respondents, in order to achieve 40-50 observations per health state. Given an absence of guidance on sample size requirements for such studies, reference to the literature in this area indicated this target sample size was appropriate. To ensure that the sample was representative of the UK general population, rigid minimum quotas were set for age group (in ten-year age bands), gender and socio-economic group. Ethical approval for both surveys was obtained from the University of Exeter Medical School Research Ethics Committee.

The labelling of MS health states (Survey 2) was informed by the approach taken in a similar previous study[7]. The condition was mentioned in the introduction to the survey and in the health state descriptions that were used in the valuation tasks. Respondents were provided

with information about the condition, based on the definition provided on the UK NHS Choices website (Figure 2)[22] prior to completing the TTO questions. A link to the MS pages of the NHS Choices website was also included. Respondents to the labelled version of the survey were asked which, if any, sources of information about MS they used to assist them in completing the TTO tasks (used survey description, they already know about MS, looked up additional information). One of the dimensions of the MSIS-8D specifically referred to 'your MS'. In order to remove any mention of MS from the unlabelled version of the survey, the phrase 'your MS' was replaced with 'your health'. This amended wording was retained for the labelled version of the survey, but the description of the health state was preceded by the phrase 'Due to having MS'.

< Insert Figure 2 >

The procedure for valuing states considered worse than being dead was consistent with the MVH approach. Prior to analysis, a monotonic transformation was applied to transform negative values onto a scale from 0 to -1[14]. Data was excluded from respondents who provided responses that appeared internally inconsistent or illogical, specifically those where respondents:

i. gave the same value to all six health states (unless they valued all health states as equivalent to full health),
ii. gave all states a value less than or equal to zero,
iii. valued the pits state at least as highly as all other states,
iv. gave the least severe state a lower value than all other states, or
v. provided three or more inconsistent responses with a difference in HSV of at least 0.1 i.e. they valued a dominated health state as better than a logically better alternative by the equivalent of one year in the TTO exercise.

Analysis of the impact of adding an MS label to MSIS-8D health states was undertaken in two phases. In the first phase, mean HSVs from the labelled and unlabelled versions of the MSIS-8D valuation survey were compared using two-sided *t*-tests, and comparisons made by severity group. In addition, data from both surveys was combined into a single dataset and a regression model was estimated using the following specification (Rowen et al, 2012)[7]:

$y_{ij} = \alpha + f(\beta' X_{\lambda\partial}) + \gamma q_i + \varepsilon_{ij}$

$i = 1, 2, \ldots, n$: individual health states
$j = 1, 2, \ldots, m$: individual respondents
y_{ij}: health state i valued by respondent j
X: vector of dummy variables for each level λ of dimension ∂ of the classification system, where level λ = 1 acts as a baseline for each dimension;
q: dummy variable for labelling effects
ε_{ij}: error term.

This model, using random effects and a generalised least squares structure in order to allow for multiple observations per respondent[23], aims to determine whether inclusion of an MS label influenced HSVs when these were modelled using dimension-levels for each health state description. Model performance was assessed using root mean squared error, *R*-squared statistics and the Wald chi-squared test[7].

In the second phase of the analysis, random effects models were estimated to produce a full tariff of HSVs for the MSIS-8D descriptive system (i.e. CSPBM) from each version of the survey, as follows[12]:

$y_{ij} = f(\beta' X_{\lambda\partial}) + \varepsilon_{ij}$

In order to ensure comparability between models, the same model specification was used for both labelled and unlabelled data[16]. Where a lower level of severity resulted in a greater decrement (higher negative coefficient) to the HSV than a higher level of severity, levels were merged and the analysis was re-run to ensure a consistent model[24].

Models were compared in terms of their performance and predictive ability, using the following criteria:

i. the proportion of coefficients that are statistically significant;
ii. the number of coefficients that are not consistent with the dimension levels;
iii. mean absolute error (MAE);
iv. proportion of health states with prediction errors > 0.05 and > 0.10;
v. R-squared statistics.

The size and significance of coefficients was compared to investigate differences in the weighting of individual dimensions and in the effect of moving between dimension-levels, and the ranges of HSVs predicted by the models were compared. In order to simplify the comparison between the coefficients of the two models, a rule of thumb was applied such that any absolute difference less than 0.01 would be considered negligible.

Financial support for this study was provided in part by the Multiple Sclerosis Society of Great Britain and Northern Ireland. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

3. Results

The unlabelled survey (Survey 1) was completed by 1702 respondents, and 1788 completed the labelled survey (Survey 2). A slightly higher percentage of respondents were excluded from the labelled survey (8.2%) compared to the unlabelled survey (7.4%) due to inconsistent or illogical responses. Most of the exclusions (>75%) were due to respondents valuing the pits (worst) state at least as highly as all other states. Data from 1576 respondents to the unlabelled survey and 1641 respondents to the labelled survey were included in the analysis. The median number of observations per health state was 47 for the unlabelled survey and 49 for the labelled survey.

Table 1 reports the demographic characteristics of participants compared to the UK population. Overall, the samples were reasonably representative of the UK population. Although the percentage of participants aged over 65 was similar to the general population, both samples included a higher proportion from the 66-75 year age band and fewer respondents aged over 75. A lower proportion of the samples was in employment or had no qualifications. There were no differences in the socio-demographic profiles of the two samples.

< Insert Table 1 >

The mean directly elicited values for each health state in the unlabelled survey ranged from 0.08 for the pits state to 0.89 for the best health state described by the MSIS-8D, and from 0.07 to 0.89 for the labelled survey (see Appendix 1). Table 2 reports data on understanding of the TTO questions and participant level of ease or difficulty in making choices. There was little difference between the labelled and unlabelled versions of the survey. A large proportion of respondents (81.4%) reported that they used the description of MS that was provided in the survey. Around one third (34.7%) reported that they "already knew quite a bit about MS". Only 6.1% reported that they had looked up additional information about MS. Completion times were similar across surveys with a mean completion time of 10 minutes 25 seconds (sd: 8 mins 44 seconds).

< Insert Table 2 >

Table 3 reports comparison of mean HSVs, showing a significant difference in overall mean HSVs between the labelled and unlabelled versions of the survey. The mean HSV was 0.023 lower when the MS label was included in the health state descriptions. Mean HSVs for the severity groups reflect the expected direction of preferences, with HSVs decreasing as severity increases. The difference in means was statistically significant ($p < 0.05$) for two

severity groups: the mildest group (Group 1) and one of the moderate groups (Group 4). At the 10% level, the difference between means was significant for the four mildest severity groups. The differences between the means for the most severe group (Group 5) and for the pits state were not statistically significant.

< Insert Table 3 >

The results of the model estimated using data from both samples are presented in Table 4. Results broadly reflect the expected direction of preferences, with mean HSVs decreasing as severity increases. The impact of the condition label is statistically significant, with a negative coefficient, indicating that the presence of an MS label lowers mean HSVs. The *p*-value for the Wald chi-squared test was <0.001 .

< Insert Table 4 >

Table 5 reports the sets of full CSPBM tariffs, across all described health states, estimated from the labelled and unlabelled datasets (i.e. MSIS-8D-unlabelled, and MSIS-8D-labelled). The coefficients of both models had the expected negative sign, i.e. as impairments in health status worsened for each dimension this has a negative effect on HSVs. All model coefficients for the unlabelled version of the survey were consistent with expected preferences, coefficients increasing as the level of impairment increased, with lower HSVs for increased impairment on that dimension. In the initial random effects model for the labelled version of the survey, three coefficients were inconsistent with the expected direction of preference. Therefore, levels were merged to create a consistent model which was re-estimated using data from the labelled version of the survey by merging levels 2 and 3 for the physical, social and emotion domains.

Predicted HSVs ranged from 0.079 to 0.883 for the unlabelled version of the survey and from 0.066 to 0.861 for the labelled version. The differences between the coefficients of the two models are illustrated in Figure 3.

< Insert Table 5 >

<Insert Figure 2>

The addition of the MS label resulted in an increase in absolute coefficient sizes for five of the MSIS-8D dimensions; an increase in coefficient size indicates that moving from level 1 to the level corresponding to the coefficient has a greater negative impact on HSVs. These were: *Physical* (coefficients for levels 2 and 3), *Social* (levels 2 and 4), *Mobility* (all levels),

Fatigue (level 3), and *Cognition* (levels 3 and 4). The labelled version had a higher number of significant coefficients for the *Mobility*, *Fatigue* and *Cognition* dimensions.

The addition of the MS label resulted in a reduction in absolute coefficient sizes for the remaining three dimensions; a decrease in coefficient size indicates that moving from level 1 to the level corresponding to the coefficient has a smaller negative impact on HSVs. These were: *Daily activities* (all levels), *Emotion* (levels 3 and 4), and *Depression* (levels 2 and 4).

Table 5 also reports the effect of including an MS label on the distance between adjacent coefficients, reflecting the impact of moving between levels on each dimension of the MSIS-8D, using the aforementioned rule of thumb (on differences >0.01). This suggests a more complex scenario than was apparent from the overall effects on coefficient sizes. For example, if we move from level 1 to level 2 on the *Physical* dimension, this has a larger impact on HSVs when modelled using labelled data, but moving from level 3 to level 4 on the same dimension has a larger impact when modelled using the unlabelled data. Indeed, four of the eight MSIS-8D dimensions have a mix of larger and smaller single-level increments when the two models are compared (*Physical*, *Social, Fatigue, Depression*).

When the condition was labelled more of the dimension coefficients were significant (n=14), compared to the model for the unlabelled data (n=11). Conversely, inclusion of the MS label was detrimental to the performance and predictive ability of the estimated model, reducing the size of the within and overall R-squared statistics, slightly increasing the MAE and RMSE and producing a greater number of health states with prediction errors greater than 0.1 or 0.05. There was no difference between the Wald chi-squared p-values of the models.

4. Discussion

Results indicate that adding an MS label to descriptions of health states significantly reduces estimated overall mean health state values, compared with values elicited for otherwise identical unlabelled states. Data from directly elicited HSVs suggests that the MS label had a more significant effect on values for less severe health states and no significant effect on values for the most severe states, reflecting a reversal of the findings reported by Rowen et al[7], who found that including a cancer label had a more significant impact on severe states.

However, this overall finding masks a more complex underlying pattern in the data, as the inclusion of the MS label had a differential impact across the dimensions of the MSIS-8D.

For mental health dimensions (depression and emotional well-being) and for daily activities, the impact of the condition label was to reduce the impact of impairments (increasing severity) on the HSVs, i.e. smaller decrements in the HSV when impairments present compared to unlabelled data. Whereas the impact of the condition label on estimated HSVs of limitations in physical health dimensions (physical and mobility), social activities, fatigue and cognition was increased, i.e. greater decrements in the HSV when impairment present compared to unlabelled data. The findings suggest interventions that target physical symptoms, social activities, fatigue and cognition may appear more effective (greater health gain if symptoms alleviated), and hence more cost-effective, if assessed using the labelled version of the MSIS-8D model, while interventions that improve mental and emotional wellbeing may appear less effective and cost-effective, compared to assessment with unlabelled data.

An added level of complexity is introduced by the differences between the models in terms of the impact on HSVs of moving between individual levels within each dimension. Furthermore, the model based on labelled health state data was more sensitive to changes in moderate levels of some domains, with more significant coefficients at level 2 (mobility) and level 3 (mobility, fatigue and cognition).

Findings suggest that the inclusion of a condition label, when estimating HSVs, may have important implications for the results of economic evaluations. In a decision analytic context it is the difference (interval) between health state values (e.g. pre- and post- intervention) that are most relevant rather than absolute HSVs. For example, if all labelled health states were valued 0.02 higher than all unlabelled health states, both sets of HSV data should produce identical values for any given change in health status and there would be no difference in the results of economic evaluations (assuming no impacts from mortality effects). It is the difference in the relative weighting of dimensions and in the effect of changes in individual dimension-levels that is relevant in determining the effects of condition labels on the results of cost effectiveness analysis. To illustrate, a shift from health state 22323222 (level 3 on the mobility and fatigue dimensions, level 2 on all other dimensions) to health state 22222222 (level 2 on all dimensions) produces an improvement in HSV of 0.049 (0.673 - 0.624) using the labelled version of the MSIS-8D tariff, compared to 0.021 (0.717 – 0.696) using the unlabelled version. Conversely, a shift from health state 22222322 (level 3 on emotion, level 2 on all other dimensions) to health state 22222221 (level 1 on depression, level 2 on all other dimensions) produces an improvement of 0.016 (0.689 - 0.673) using the labelled version and of 0.053 (0.743 – 0.689) using the unlabelled version.

There was little difference in the performance or predictive ability of the two models. The model based on labelled data had slightly lower predictive ability in terms of the R-squared statistics, MAE, RMSE and the number of health states with prediction errors greater than 0.1 or 0.05. This may be due to the increased cognitive demand placed on respondents to the labelled version of the study, who were asked to imagine that they had MS in addition to imagining themselves in particular health states.

There are a number of limitations with the study. In keeping with previous guidance[3], valuation methods were consistent with the MVH version of the TTO protocol. One consequence of this is that the protocol asks respondents to imagine remaining in a specified health state for ten years, with no changes in that health state during that time. However, the definition of MS that was provided for respondents to the labelled version of the survey stated that MS is usually characterised by alternating periods of relapse and remission, or by ongoing progression. This may have caused confusion for respondents to the labelled version of the survey and may have affected the values they attributed to health states. This is a potential area for further research.

Concerns have been raised about whether TTO tasks may pose too high a cognitive burden for online administration[25]. While recent evaluations of internet TTO have produced mixed results[26-28], the nature of the preference data elicited here, alongside respondents' self-reported task comprehension, suggest that the TTO technique can be administered successfully online.

A central limitation of this study is that it is quantitative in nature and therefore cannot explain why HSVs differ between labelled and unlabelled health states. Reasons for the differences in the health state values require investigation using qualitative methods. There are a number of possible mechanisms that may explain differences between values elicited for labelled and unlabelled health states. It may be that differences arise due to a legitimate effect of the condition *per se*, for example contextual effects that provide differing explanations for impairments in functioning. Alternatively, labelling health states may prompt respondents to take irrelevant or inaccurate factors into account, or lead to 'focussing effects', thereby distorting HSVs. Use of a condition label may lead to respondents own preconceptions about the condition being overly prominent[29], and/or increased influence of social attitudes[24], and/or poor hedonic forecasting[30]. Further research in this area is planned to explore how and why respondents' cognitive processes differ when undertaking TTO tasks for labelled and unlabelled MSIS-8D health states. This will also enable investigation of the suitability of the ten-year time-frame for a variable, progressive condition, as mentioned above.

To our knowledge, this is the first study to undertake two full-scale preference elicitation surveys for the same condition-specific classification system, using large, representative samples of the general population and a large sample of health states, in order to compare both observed and predicted values for labelled and unlabelled health states. Results presented and discussed here show that including an MS label in health state descriptions reduces the overall estimated health state values associated with MSIS-8D health states and affects the relative importance of individual dimensions of HRQoL in influencing HSVs. The differences reported here between health state values, by labelled versus unlabelled states, could be considered relatively small, in a descriptive context. However, in economic evaluations mean HSVs and related QALY differences (between comparators) are often relatively small and we consider the findings here to be important, firstly in addressing an area of methodological uncertainty, but also for applied decision making contexts. We acknowledge that it will only be in the consideration of any differences (in HSVs) in a decision-analytic context (policy, CUA setting) that we may see the magnitude of the impact of using condition labels to elicit preferences and to estimate HSVs, and this is the focus for ongoing and future research.

References

1. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value in Health. 2009; 12 Suppl 1:S5-9.
2. Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health. 2009; 12 Suppl 1:S10-5.
3. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. 2013 [Accessed 26 September 2016]Available from: http://www.nice.org.uk/article/pmg9/chapter/foreword
4. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.
5. Pharmaceutical Benefits Advisory Committee. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee. Canberra: Australian Government - Department of Health and Ageing; 2008.
6. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. Journal of the American Medical Association. 2016; 316(10):1093-103.
7. Rowen D, Brazier J, Tsuchiya A, Young T, Ibbotson R. It's all the in the name, or is it? The impact of labelling on health state values. Medical Decision Making. 2012; 32:31–40.
8. Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. Journal of Clinical Epidemiology. 1993; 46:77–84.
9. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. Describing health states. Methodologic issues in obtaining values for health states. Medical Care. 1984; 22:543–52.
10. Rabin R, Rosser RM, Butler C. Impact of diagnosis on utilities assigned to states of illness. Journal of the Royal Society of Medicine. 1993; 86:444–8.
11. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. Journal of Chronic Diseases. 1978; 31.
12. Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technology Assessment. 2012; 16:1-114.
13. Goodwin E, Green C. A systematic review of the literature on the development of condition-specific preference-based measures. Applied Health Economics and Health Policy. 2016; 14:161-83.
14. Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997; 35:1095-108.
15. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002; 21 271-92.
16. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Economics. 2010; 19:125-9.
17. Goodwin E, Green C. A QALY Measure for Multiple Sclerosis: Developing a Patient-Reported Health State Classification System for an MS-Specific Preference-Based Measure. Value in Health. 2015; 18:1016-24.
18. Hobart JC, Riazi A, Lamping D, Fitzpatrick R, Thompson AJ. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. Health Technology Assessment. 2004; 8:1-48.
19. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technology Assessment. 2009; 13:1-177.
20. Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight dimensional health state classification system derived from the Multiple Sclerosis Impact Scale (MSIS-29). Value in Health. 2015; 18:1025-36.
21. Time trade-off user manual: Props and self-completion methods. In: Gudex C, ed. York: The MVH Group, Centre for Health Economics, University of York 1994.
22. NHS Choices. Multiple sclerosis. 17/02/2016 [Accessed 26 September 2016]Available from: http://www.nhs.uk/conditions/Multiple-sclerosis/Pages/Introduction.aspx
23. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press 2007.
24. Mulhern B, Rowen D, Brazier J, Smith S, Romeo R, Tait R, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: Generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technology Assessment. 2013; 17:1-140.
25. Attema AE, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. Eur J Health Econ. 2013; 14 Suppl 1:S53-64.

26. Norman R, King MT, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual Life Res. 2010; 19(4):499-508.
27. Versteegh MM, Attema AE, Oppe M, Devlin NJ, Stolk EA. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. Eur J Health Econ. 2013; 14 Suppl 1:S43-51.
28. Mulhern B, Bansback N, Brazier J, Buckingham K, Cairns J, Devlin N, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Health Technol Assess. 2014; 18(12):vii-xxvi, 1-191.
29. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Quality of Life Research. 2003; 12:599-607.
30. Stigglebout AM, de Vogel-Voort E. Health state utilities: a framework for studying the gap between the imagined and the real. Value in Health. 2008; 11:76-87.

Figure 1: MSIS-8D classification system

Physical subscale (includes all physical and social dimensions)					
Dimension	**MSIS-29 item**	**Response levels**			
Physical	In the past two weeks, how much has your MS limited your ability to do physically demanding tasks?	1	2	3	4
	In the past two weeks, how much have you been bothered by:				
Social	Limitations in your social and leisure activities at home?	1	2	3	4
Mobility	Being stuck at home more than you would like to be?	1	2	3	4
Daily activities	Having to cut down the amount of time you spent on work or other daily activities?	1	2	3	4
Psychological subscale (includes all psychological and other non-physical dimensions)					
Dimension	**MSIS-29 item**	**Response levels**			
	In the past two weeks, how much have you been bothered by:				
Cognition	Feeling mentally fatigued?	1	2	3	4
Emotion	Feeling irritable, impatient or short-tempered?	1	2	3	4
Cognition	Problems concentrating?	1	2	3	4
Depression	Feeling depressed?	1	2	3	4
Response levels: 1 = not at all; 2 = a little; 3 = moderately; 4 = extremely					

Figure 2: Information about MS provided for respondents to the labelled survey

In the following questions, we would like you to imagine that you have multiple sclerosis. Multiple sclerosis affects nerves in the brain and spinal cord, causing a wide range of symptoms including:

- loss of vision – usually only in one eye
- spasticity – muscle stiffness that can lead to uncontrolled muscle movements
- ataxia – difficulties with balance and co-ordination
- fatigue – feeling very tired during the day

Around eight out of 10 people with multiple sclerosis (MS) are diagnosed with the relapsing remitting type of MS. Someone with relapsing remitting MS will have flare-ups of symptoms, known as relapses. These can last from a few days to a few months. These will be followed by periods where symptoms are mild or disappear altogether. This is known as remission and can last for days, weeks or sometimes months.

Usually after around 15 years, around half of people with relapsing remitting MS will go on to develop secondary progressive MS. In secondary progressive MS, symptoms gradually worsen over time. Some people may still have relapses, but without full recovery from symptoms.

The least common form of MS is primary progressive MS. In this type, symptoms gradually get worse over time and there are no periods of remission.

Taken from the NHS Choices website, see:

http://www.nhs.uk/conditions/Multiple-sclerosis/Pages/Introduction.aspx

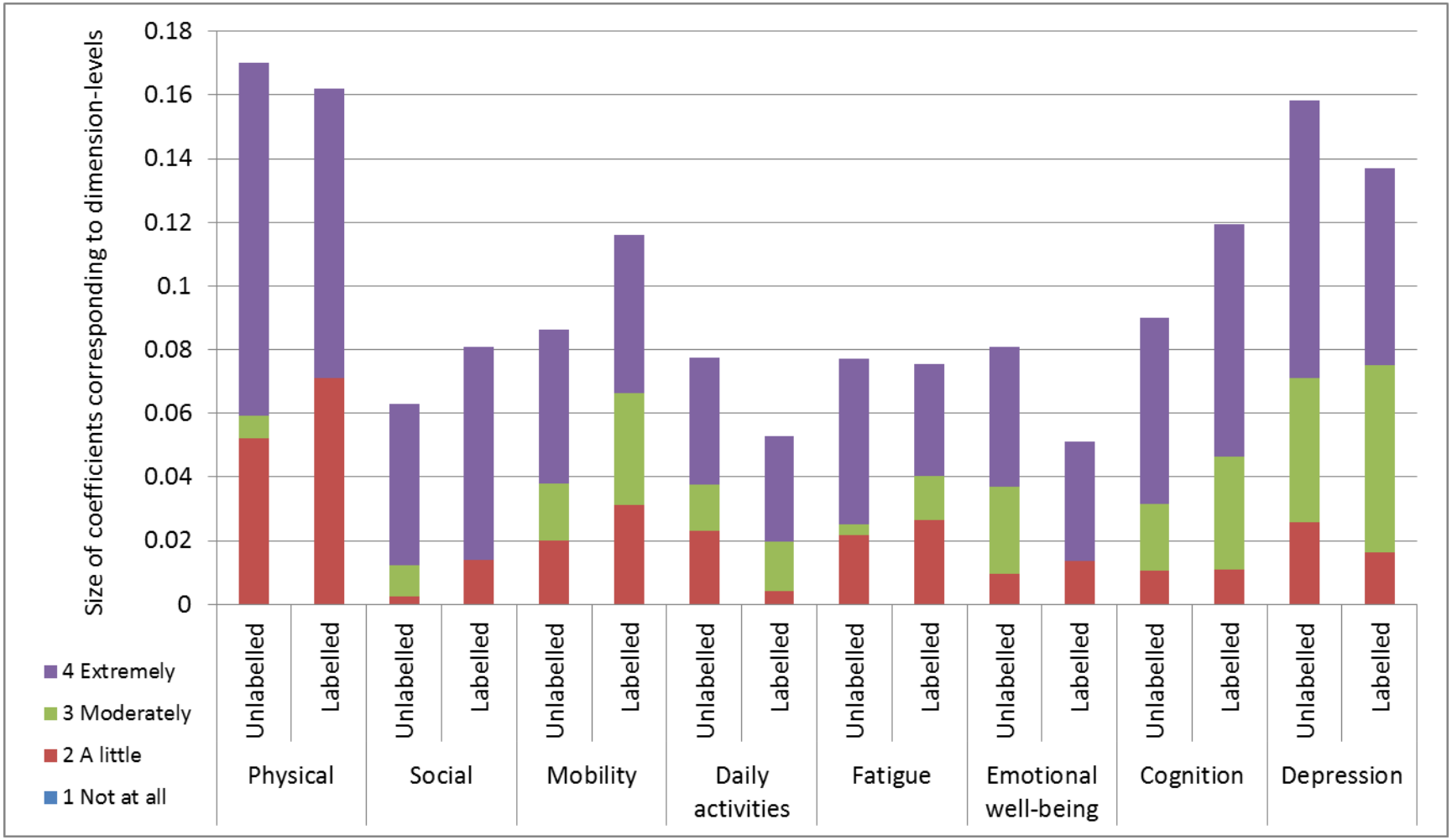

Figure 3: Comparing the size of coefficients by dimension and level for the valuation models of labelled and unlabelled health state

Table 1: Characteristics of respondents

	Unlabelled survey (n=1576)		Labelled survey (n = 1641)		UK
Nationality					
England	1378	87.44%	1387	84.52%	83.84%
Northern Ireland	23	1.46%	29	1.77%	2.77%
Scotland	117	7.42%	125	7.62%	8.52%
Wales	58	3.68%	100	6.09%	4.87%
English region (n = 1387)					
East Midlands	100	7.26%	119	8.58%	8.58%
East of England	135	9.80%	144	10.38%	11.03%
Greater London	173	12.55%	153	11.03%	15.34%
North East England	73	5.30%	76	5.48%	4.94%
North West England	192	13.93%	189	13.63%	13.26%
South East England	276	20.03%	269	19.39%	16.28%
South West England	154	11.18%	166	11.97%	10.16%
West Midlands	138	10.01%	115	8.29%	10.45%
Yorkshire and the Humber	137	9.94%	156	11.25%	9.94%
Gender					
Female	829	52.60%	865	52.71%	51.40%
Male	747	47.40%	776	47.29%	48.60%
Age group					
18-25	206	13.07%	197	12.00%	13.56%
26-35	269	17.07%	278	16.94%	16.88%
36-45	261	16.56%	256	15.60%	17.48%
46-55	279	17.70%	297	18.10%	17.45%
56-65	253	16.05%	293	17.85%	14.69%
66-75	279	17.70%	285	17.37%	10.83%
76-85	26	1.65%	35	2.13%	6.69%
86 or over	3	0.19%	0	0.00%	2.41%
Employment status					
Employed	801	50.82%	817	49.79%	60.05%
Economically inactive	681	43.21%	738	44.97%	35.50%
Unemployed/ seeking work	89	5.65%	76	4.63%	4.45%
Other/ Prefer not to say	5	0.32%	10	0.61%	
Highest level of education					
Level 4	472	29.95%	497	30.29%	27.02%
Level 3	459	29.12%	517	31.51%	12.12%
Level 1 or 2	443	28.11%	435	26.51%	29.24%
Trade apprenticeships	61	3.87%	60	3.66%	3.30%
Other qualifications	46	2.92%	51	3.11%	5.13%
No qualifications	89	5.65%	78	4.75%	23.19%
Prefer not to say	6	0.38%	3	0.18%	
Socio-economic group					
AB	422	26.78%	462	28.15%	25%
C1	495	31.41%	469	28.58%	29%
C2	295	18.72%	350	21.33%	21%
DE	364	23.10%	360	21.94%	25%

Sources for UK data:
Nationality, region, gender and age group: ONS mid-2012 population estimates, age 18 and over
Employment status: ONS labour market statistics, August to December 2013, age 18 and over
Educational attainment: 2011 UK Census, age 16 and over
Socio-economic group: provided by Accent Marketing and Research Ltd

Table 2: Self-reported task comprehension

What were the questions like to understand?					How easy or difficult was it to make choices?				
	Unlabelled		Labelled			Unlabelled		Labelled	
Very easy	617	39.15%	676	41.19%	Very easy	250	15.86%	288	17.55%
Easy	825	52.35%	816	49.73%	Easy	583	36.99%	538	32.78%
Difficult	126	7.99%	137	8.35%	Difficult	668	42.39%	702	42.78%
Very difficult	8	0.51%	12	0.73%	Very difficult	75	4.76%	113	6.89%

Table 3: Comparison between labelled and unlabelled data, by severity group

		Unlabelled	Labelled	Difference	t stat	Pr(T > t)
All health states	Mean HSV	0.484	0.461	0.023	3.397	0.0007
	SD	0.0048	0.0049			
Severity Group 1	Mean HSV	0.8173	0.7803	0.0370	4.1528	<0.0001
(least severe)	SD	0.2330	0.2697			
Severity Group 2	Mean HSV	0.6818	0.6654	0.0164	1.4143	0.0787
	SD	0.3193	0.3364			
Severity Group 3	Mean HSV	0.5782	0.5586	0.0196	1.4314	0.0762
	SD	0.3830	0.3939			
Severity Group 4	Mean HSV	0.4733	0.4393	0.0340	2.1983	0.0140
	SD	0.4213	0.4540			
Severity Group 5	Mean HSV	0.2693	0.2483	0.0210	1.2038	0.1144
	SD	0.4828	0.5042			
Pits health state	Mean HSV	0.0830	0.0715	0.0115	0.6676	0.2522
	SD	0.4800	0.4937			

Number of observations in each severity group (and for the pits state): unlabelled = 1576; labelled = 1641; combined = 3217. Degrees of freedom = 3215. SD = standard deviation

Table 4: Results of initial regression analysis

Dimension and level		Coefficient	P>z
Physical	2	-0.062	0.000
	3	-0.064	0.000
	4	-0.164	0.000
Social	2	-0.009	0.431
	3	-0.014	0.423
	4	-0.073	0.000
Mobility	2	-0.027	0.020
	3	-0.053	0.002
	4	-0.101	0.000
Daily activities	2	-0.013	0.275
	3	-0.028	0.094
	4	-0.065	0.001
Fatigue	2	-0.024	0.021
	3	-0.035	0.026
	4	-0.079	0.000
Emotion	2	-0.011	0.285
	3	-0.021	0.174
	4	-0.062	0.003
Cognition	2	-0.011	0.321
	3	-0.041	0.012
	4	-0.107	0.000
Depression	2	-0.021	0.060
	3	-0.073	0.000
	4	-0.147	0.000
MS label		-0.024	0.033
Constant		0.884	0.000
Performance			
sigma_u			0.293
sigma_e			0.277
rho			0.529
RMSE			0.277
R-squared			
within			0.496
between			0.010
overall			0.282
Wald chi^2			15825.34
Prob chi^2			<0.001
Observations = 19302; groups = 3217 Base case: Dimension Level 1, unlabelled			

Table 5: Impact of MS label on estimated HSV tariffs for MSIS-8D

Dimension	Level	Model coefficients			Impact of moving between levels			
		Unlabelled	Labelled	Diff		Unlabelled	Labelled	Diff
Physical	2	-0.052**	-0.071**	+0.019	1 to 2	0.052	0.071	+0.019
	3	-0.059*		+0.012	2 to 3	0.007	0.000	-0.007
	4	-0.170**	-0.162**	-0.008	3 to 4	0.111	0.091	-0.020
Social	2	-0.002	-0.014	+0.012	1 to 2	0.002	0.014	+0.012
	3	-0.012		+0.002	2 to 3	0.010	0.000	-0.010
	4	-0.063*	-0.081**	+0.018	3 to 4	0.051	0.067	+0.016
Mobility	2	-0.020	-0.031*	+0.011	1 to 2	0.020	0.031	+0.011
	3	-0.038	-0.066**	+0.028	2 to 3	0.018	0.035	+0.017
	4	-0.086**	-0.116**	+0.030	3 to 4	0.048	0.049	+0.001
Daily activities	2	-0.023	-0.004	-0.019	1 to 2	0.023	0.004	-0.019
	3	-0.038	-0.020	-0.018	2 to 3	0.015	0.015	+0.001
	4	-0.077**	-0.053*	-0.025	3 to 4	0.040	0.033	-0.007
Fatigue	2	-0.022	-0.027	+0.005	1 to 2	0.022	0.027	+0.005
	3	-0.025	-0.040*	+0.015	2 to 3	0.003	0.014	+0.011
	4	-0.077**	-0.076**	-0.002	3 to 4	0.052	0.035	-0.017
Emotion	2	-0.009	-0.014	+0.004	1 to 2	0.009	0.014	+0.004
	3	-0.037		-0.023	2 to 3	0.027	0.000	-0.027
	4	-0.081**	-0.051*	-0.030	3 to 4	0.044	0.037	-0.007
Cognition	2	-0.011	-0.011	0.000	1 to 2	0.011	0.011	0.000
	3	-0.032	-0.046*	+0.015	2 to 3	0.021	0.035	+0.014
	4	-0.090**	-0.119**	+0.029	3 to 4	0.058	0.073	+0.015
Depression	2	-0.026	-0.016	-0.010	1 to 2	0.026	0.016	-0.010
	3	-0.071**	-0.075**	+0.004	2 to 3	0.045	0.059	+0.014
	4	-0.158**	-0.137**	-0.021	3 to 4	0.087	0.062	-0.025
Constant		0.883**	0.861**	-0.022				

Model performance	**Unlabelled**	**Labelled**
Coefficients:	24	21
significant (p<0.05)	45.83% (11)	66.67% (14)
significant (p<0.1)	45.83% (11)	71.43% (15)
R-squared: within	0.5028	0.4894
between	0.0068	0.0109
overall	0.2953	0.2688
Wald chi^2 (prob > chi^2)	7959.82 (<0.001)	7864.32 (<0.001)
RMSE	0.276	0.278
MAE	0.041	0.044
Errors > 0.1	9 (5.33%)	11 (6.51%)
Errors > 0.05	49 (28.99%)	65 (38.46%)
Range of HSVs	0.079 to 0.883	0.066 to 0.861
Observations (groups)	9456 (1576)	9846 (1641)

* p<0.05; **p<0.01. Levels 2 and 3 were merged in the labelled version for Physical, Social and Emotion
Diff = difference:
+ indicates that the MS label increased the size of the coefficient or the impact of moving between levels
- indicates that the MS label reduced the size of the coefficient or the impact of moving between levels

Appendix 1: Mean (SD) for directly elicited health state values

	Unlabelled survey			Labelled survey		
Health state	**Mean HSV**	**SD**	**Obs**	**Mean HSV**	**SD**	**Obs**
11111111	0.893	0.177	50	0.893	0.260	40
11111112	0.870	0.179	51	0.826	0.222	46
11111121	0.872	0.231	45	0.846	0.266	50
11111211	0.855	0.209	93	0.825	0.246	93
11112111	0.905	0.169	47	0.825	0.260	41
11121111	0.872	0.202	38	0.825	0.265	67
11211111	0.848	0.209	43	0.871	0.195	57
12111111	0.873	0.221	53	0.818	0.239	48
21111111	0.829	0.259	96	0.794	0.286	90
11112211	0.868	0.195	47	0.852	0.227	50
21112111	0.743	0.260	44	0.764	0.251	50
21121111	0.845	0.199	46	0.825	0.216	50
11112221	0.876	0.234	51	0.808	0.210	49
21112211	0.808	0.214	47	0.713	0.337	41
21122111	0.788	0.241	49	0.748	0.285	50
22121111	0.852	0.206	38	0.779	0.231	58
11112222	0.783	0.226	49	0.775	0.249	49
21112221	0.827	0.174	37	0.720	0.400	49
21122211	0.822	0.250	49	0.796	0.245	50
22122111	0.810	0.210	50	0.735	0.299	62
32121111	0.839	0.239	44	0.756	0.266	51
11113222	0.787	0.231	55	0.763	0.247	51
21112222	0.720	0.249	50	0.773	0.244	46
21122221	0.733	0.260	41	0.777	0.296	46
22122211	0.733	0.284	40	0.766	0.244	45
32122111	0.796	0.230	50	0.769	0.273	40
32221111	0.816	0.244	45	0.726	0.310	43
11113322	0.799	0.214	51	0.794	0.240	50
21113222	0.748	0.228	39	0.722	0.252	52
21122222	0.739	0.299	49	0.715	0.289	40
22122221	0.705	0.246	41	0.733	0.380	46
32122211	0.751	0.333	41	0.614	0.275	43
32222111	0.782	0.188	48	0.656	0.336	44
42221111	0.612	0.282	43	0.632	0.352	49
11113332	0.777	0.218	37	0.762	0.289	49
21113322	0.744	0.279	47	0.715	0.295	50
21123222	0.759	0.332	38	0.718	0.300	67
22122222	0.671	0.332	40	0.772	0.237	45
32122221	0.762	0.225	43	0.804	0.226	57
32222211	0.727	0.266	50	0.762	0.293	40
42222111	0.649	0.348	39	0.725	0.237	52
42231111	0.628	0.383	49	0.657	0.384	50
11114332	0.769	0.286	47	0.627	0.332	41
21113332	0.694	0.290	52	0.654	0.387	39

	Unlabelled survey			Labelled survey		
Health state	**Mean HSV**	**SD**	**Obs**	**Mean HSV**	**SD**	**Obs**
21123322	0.751	0.339	44	0.642	0.368	51
22123222	0.721	0.288	50	0.708	0.245	62
32122222	0.755	0.294	49	0.748	0.208	50
32222221	0.722	0.338	51	0.752	0.235	49
42222211	0.638	0.401	53	0.601	0.378	48
42232111	0.650	0.279	49	0.634	0.353	49
42331111	0.589	0.384	50	0.620	0.381	40
11114333	0.706	0.245	41	0.560	0.407	43
21114332	0.612	0.379	44	0.609	0.417	50
21123332	0.636	0.401	45	0.680	0.261	50
22123322	0.741	0.268	51	0.667	0.301	46
32123222	0.730	0.248	46	0.727	0.326	50
32222222	0.647	0.319	49	0.686	0.312	40
42222221	0.720	0.279	38	0.571	0.403	58
42232211	0.650	0.375	45	0.646	0.356	43
42332111	0.611	0.383	51	0.559	0.443	50
43331111	0.542	0.431	50	0.600	0.396	44
21114333	0.595	0.324	43	0.576	0.347	49
21124332	0.688	0.200	48	0.561	0.303	44
22123332	0.690	0.242	50	0.683	0.306	46
32123322	0.693	0.228	44	0.648	0.334	51
32223222	0.723	0.244	55	0.718	0.285	51
42222222	0.505	0.431	41	0.611	0.381	46
42232221	0.616	0.416	55	0.511	0.383	51
42332211	0.652	0.250	48	0.514	0.371	44
43332111	0.580	0.334	45	0.516	0.459	50
43341111	0.640	0.274	47	0.594	0.317	41
21124333	0.610	0.363	43	0.668	0.348	57
22124332	0.615	0.313	46	0.685	0.338	50
32123332	0.620	0.444	41	0.588	0.357	43
32223322	0.702	0.285	38	0.653	0.343	67
42223222	0.664	0.299	49	0.600	0.329	50
42232222	0.661	0.307	47	0.621	0.339	50
42332221	0.563	0.417	50	0.589	0.397	40
43332211	0.577	0.401	49	0.531	0.420	40
43342111	0.500	0.436	53	0.510	0.427	48
43441111	0.529	0.387	50	0.482	0.408	44
21124343	0.450	0.455	52	0.479	0.474	39
22124333	0.500	0.449	44	0.539	0.415	50
32124332	0.688	0.401	51	0.631	0.288	49
32223332	0.754	0.247	44	0.596	0.350	51
42223322	0.552	0.355	45	0.598	0.337	43
42233222	0.543	0.384	44	0.524	0.402	51
42332222	0.576	0.355	37	0.448	0.536	49
43332221	0.637	0.334	47	0.470	0.441	41

	Unlabelled survey			Labelled survey		
Health state	**Mean HSV**	**SD**	**Obs**	**Mean HSV**	**SD**	**Obs**
43342211	0.543	0.433	49	0.532	0.465	50
43442111	0.520	0.402	51	0.423	0.490	46
44441111	0.622	0.356	47	0.379	0.452	41
22124343	0.510	0.461	49	0.458	0.514	50
32124333	0.553	0.345	50	0.598	0.425	40
32224332	0.530	0.421	40	0.667	0.274	45
42223332	0.514	0.372	45	0.474	0.388	50
42233322	0.521	0.416	39	0.595	0.319	52
42333222	0.526	0.404	51	0.538	0.409	50
43332222	0.564	0.391	50	0.603	0.323	62
43342221	0.449	0.511	49	0.580	0.343	49
43442211	0.596	0.364	38	0.422	0.424	58
44442111	0.467	0.437	50	0.392	0.422	46
32124343	0.506	0.430	39	0.564	0.364	52
32224333	0.604	0.356	50	0.473	0.470	44
42224332	0.455	0.410	47	0.438	0.409	41
42233332	0.565	0.330	50	0.535	0.356	62
42333322	0.606	0.438	51	0.553	0.363	49
43333222	0.532	0.417	41	0.425	0.451	43
43342222	0.644	0.311	47	0.437	0.468	50
43442221	0.441	0.467	50	0.586	0.332	40
44442211	0.381	0.504	50	0.374	0.542	40
32224343	0.572	0.254	38	0.454	0.405	58
42224333	0.442	0.404	51	0.399	0.469	46
42234332	0.452	0.446	49	0.489	0.415	40
42333332	0.580	0.378	49	0.527	0.392	50
43333322	0.410	0.419	41	0.590	0.385	46
43343222	0.616	0.332	47	0.313	0.501	41
43442222	0.484	0.364	44	0.448	0.434	51
44442221	0.337	0.479	40	0.456	0.415	45
32224443	0.407	0.457	55	0.391	0.459	51
42224343	0.456	0.363	48	0.317	0.470	44
42234333	0.435	0.427	46	0.478	0.446	50
42334332	0.418	0.388	43	0.485	0.437	49
43333332	0.479	0.450	45	0.438	0.471	43
43343322	0.417	0.509	44	0.436	0.430	50
43443222	0.447	0.359	50	0.449	0.443	46
44442222	0.380	0.478	52	0.320	0.585	39
32224444	0.351	0.487	53	0.282	0.513	48
42224443	0.347	0.385	37	0.230	0.508	49
42234343	0.493	0.445	44	0.322	0.462	51
42334333	0.440	0.419	49	0.426	0.463	49
43334332	0.479	0.422	51	0.477	0.465	50
43343332	0.512	0.411	43	0.564	0.509	57

	Unlabelled survey			Labelled survey		
Health state	**Mean HSV**	**SD**	**Obs**	**Mean HSV**	**SD**	**Obs**
43443322	0.505	0.480	43	0.565	0.434	57
44443222	0.353	0.467	38	0.301	0.480	67
42224444	0.272	0.540	51	0.266	0.585	50
42234443	0.255	0.456	44	0.298	0.451	50
42334343	0.341	0.487	50	0.393	0.482	40
43334333	0.391	0.427	52	0.370	0.538	39
43344332	0.445	0.405	51	0.394	0.502	46
43443332	0.492	0.492	44	0.360	0.461	51
44443322	0.417	0.429	46	0.475	0.444	50
42234444	0.098	0.525	55	0.190	0.515	51
42334443	0.328	0.367	37	0.204	0.510	49
43334343	0.349	0.480	53	0.260	0.478	48
43344333	0.416	0.352	48	0.280	0.476	44
43444332	0.389	0.453	45	0.410	0.433	43
44443332	0.436	0.394	47	0.203	0.520	41
42334444	0.228	0.520	38	0.125	0.497	67
43334443	0.230	0.516	41	0.404	0.474	46
43344343	0.316	0.472	45	0.333	0.488	50
43444333	0.290	0.515	41	0.181	0.514	43
44444332	0.299	0.509	39	0.322	0.484	52
43334444	0.230	0.534	49	0.195	0.525	50
43344443	0.216	0.390	50	0.273	0.565	40
43444343	0.242	0.471	47	0.168	0.485	41
44444333	0.189	0.505	40	0.327	0.523	45
43344444	0.165	0.434	50	0.190	0.479	46
43444443	0.284	0.476	50	0.203	0.528	44
44444343	0.239	0.514	47	0.257	0.492	50
34444444	0.208	0.467	38	0.084	0.474	58
43444444	0.059	0.465	43	0.124	0.371	49
44344444	0.133	0.510	49	0.119	0.510	40
44434444	0.193	0.485	50	0.135	0.487	62
44443444	0.140	0.534	51	0.175	0.514	49
44444344	0.136	0.477	49	0.039	0.512	50
44444434	0.097	0.387	44	0.194	0.482	51
44444443	0.150	0.529	49	0.018	0.530	49
44444444	0.083	0.480	1576	0.072	0.494	1641