



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Health Economics

journal homepage: [www.elsevier.com/locate/econbase](http://www.elsevier.com/locate/econbase)

## The relative value of different QALY types

Emily Lancsar<sup>a,\*</sup>, Yuanyuan Gu<sup>b</sup>, Dorte Gyrd-Hansen<sup>c</sup>, Jim Butler<sup>d</sup>, Julie Ratcliffe<sup>e</sup>, Liliana Bulfone<sup>f</sup>, Cam Donaldson<sup>g</sup><sup>a</sup> Department of Health Services Research and Policy, Research School of Population Health, Australian National University, Australia<sup>b</sup> Centre for the Health Economy, Macquarie University, Australia<sup>c</sup> Centre of Health Economics Research, Department of Public Health, University of Southern Denmark, Denmark<sup>d</sup> Health Research Institute, University of Canberra, Australia<sup>e</sup> Health and Social Care Economics Group, College of Nursing and Health Sciences, Flinders University, Australia<sup>f</sup> Deakin Health Economics, Deakin University, Australia<sup>g</sup> Yunus Centre for Social Business and Health, Glasgow Caledonian University, United Kingdom

## ARTICLE INFO

## Article history:

Received 24 October 2018

Received in revised form 26 January 2020

Accepted 29 January 2020

Available online 13 February 2020

## JEL classification:

C35

I1

I18

## Keywords:

QALYs

QALY types

Relative weights

Priority setting

Equity

Discrete choice experiments

## ABSTRACT

The oft-applied assumption in the use of Quality Adjusted Life Years (QALYs) in economic evaluation, that all QALYs are valued equally, has been questioned from the outset. The literature has focused on differential values of a QALY based on equity considerations such as the characteristics of the beneficiaries of the QALYs. However, a key characteristic which may affect the value of a QALY is the type of QALY itself. QALY gains can be generated purely by gains in survival, purely by improvements in quality of life, or by changes in both. Using a discrete choice experiment and a new methodological approach to the derivation of relative weights, we undertake the first direct and systematic exploration of the relative weight accorded different QALY types and do so in the presence of equity considerations; age and severity. Results provide new evidence against the normative starting point that all QALYs are valued equally.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

An internationally relevant policy question is how best to set priorities in the allocation of scarce public resources. Health care decision makers must make decisions about which services to fund and those not to fund. Economic evaluation plays a key role in aiding such decisions and, within this, the quality adjusted life year (QALY) is the dom-

inant measure of health gain in many countries. QALYs are computed based on individual preferences for health outcomes (under strict assumptions) (Pliskin et al., 1980; Miyamoto and Eraker, 1989). In contrast, a social decision making perspective is relevant regarding the distribution of QALYs resulting from potential resource allocation decisions. The application of the cost per QALY framework has generally focused on health maximisation (Brazier et al., 2017). While not a stated characteristic of the QALY model, in practice QALYs are often used with an (at least implicit) assumption that all QALYs are valued equally; “a QALY is a QALY” made famous by Weinstein (1988) when discussing research which questioned this basic assump-

\* Corresponding author at: Building 63, Eggleston Road, Acton, ACT 2061, Australia.

E-mail address: [Emily.Lancsar@anu.edu.au](mailto:Emily.Lancsar@anu.edu.au) (E. Lancsar).

tion (Donaldson et al., 1988; Donaldson and Wright, 1990; Weinstein, 1988). Since then, there has been interest in some prominent settings in accounting for wider social values in the context of health care resource allocation (Australian Department of Health, 2017; NICE, 2008).

Important empirical literature questioning whether all QALYs are indeed valued equally, has focused on relative weights for the characteristics of the beneficiaries of the QALY gains, arguing that the use of QALYs in economic evaluation fails to account for such contextual issues (e.g. Baker et al., 2010; Dolan and Tsuchiya, 2005; Lancsar et al., 2011; Nord and Johansen, 2014; Shah, 2009; van de Wetering et al., 2015; Wagstaff, 1991; Williams, 1997). Indeed, in a recent review of that literature, Gu et al. (2015) found that the relative social value of a QALY potentially differs according to key characteristics of the individual such as age, severity, culpability, and socio-economic status.

Recently, empirical literature has explored public preferences in relation to guidance provided by some health technology assessment (HTA) agencies indicating that the lives of patients at end-of-life (in the specific context in the case of the UK of premature death for which short extensions in survival are now possible) are essentially valued more than other patients by those agencies (Chalkidou, 2012). Here the empirical evidence is more mixed, but with a general lack of social preference for an end-of-life premium (Gu et al., 2015; Gyrd-Hansen, 2018; Shah, 2018).

A key, and to date largely under explored, way in which the social value of a QALY can differ is in relation to the type of QALY itself. While the generation of utility weights used in the calculation of QALYs (e.g. using standard gamble, time tradeoff and other methods) is based on individual preferences, this ignores potential differences in social value placed on different types of QALYs. Given a QALY is a composite figure generated by changes in length and/or quality of life, QALY gains can be generated in different ways. For example, QALYs can be primarily generated via improvements in quality of life (QoL), by extensions in survival and combinations of the two. This raises the question of whether, for example, QALYs generated by improvements in both components are valued differently by society to QALYs generated by improvements in a single component.

While the literature on generating relative weights for QALYs has not explored the relative weight accorded to different QALY types, an important related literature has calculated different monetary values for different types of QALY gain. Mason et al. (2009) indirectly modelled the value of a QoL-enhancing QALY, a life-extending QALY and a life-saving QALY from existing data on the value of a prevented statistical fatality and values of statistical injuries used in UK public sector decision making. Pennington et al. (2015) directly elicited monetary values for QoL-enhancing, life-extending and end of life QALYs for the specific case of 1-QALY gains using contingent valuation. These studies found higher monetary value associated with life extending gains over QoL-enhancing gains. They did not explore other QALY types. In both studies, monetary valuations were based on individual and not social preferences, were for a narrow range of QALY types and did

not allow trade off or interactions between QALY types and other attributes.

In the present study, we seek to directly and systematically explore whether the social value of a QALY is the same regardless of how it is generated across four QALY types (life-extending QALYs; QoL-enhancing QALYs; QALYs generated as a mix of life extension and QoL enhancement; and QALYs that extend life but simultaneously reduce QoL) and, if not, the magnitude of the strength of social preferences across such QALY types.

We do this by undertaking a discrete choice experiment (DCE) with a nationally representative sample in age and gender to explore the Australian public's preferences for which factors should receive additional weight in priority setting and what weight they should receive. Results from the estimated choice models were used to calculate relative weights for QALYs via extensions of the Hicksian compensating variation approach (Lancsar et al., 2011). Conceptually, the resulting weights may then be incorporated into HTA decision making by weighting QALYs in the numerator of an incremental cost-per-QALY ratio, or equivalently by weighting up or down a monetary threshold for a QALY (Bobinac et al., 2012; Round and Paulden, 2018). Alternatively, the weights could be used to inform discursive deliberations by appraisal committees in HTA agencies (Cookson et al., 2009; Culyer, 2006).

In addition to undertaking the first direct and systematic exploration of the relative weight accorded to different QALY types, we also add to the existing distributional literature by calculating relative priority weights across characteristics of the beneficiaries of the health gain including age and severity. An advantage of our study is that we bring these two topics together by exploring the trade-offs and interactions between QALY type and such distributional considerations all in one empirical framework.

We also make a number of methodological advancements in survey design, experimental design and calculation of relative QALY weights which collectively lead to more robust results and also offer a template to those considering calculation of relative QALY weights, or indeed other relative weights. We highlight such contributions throughout the paper.

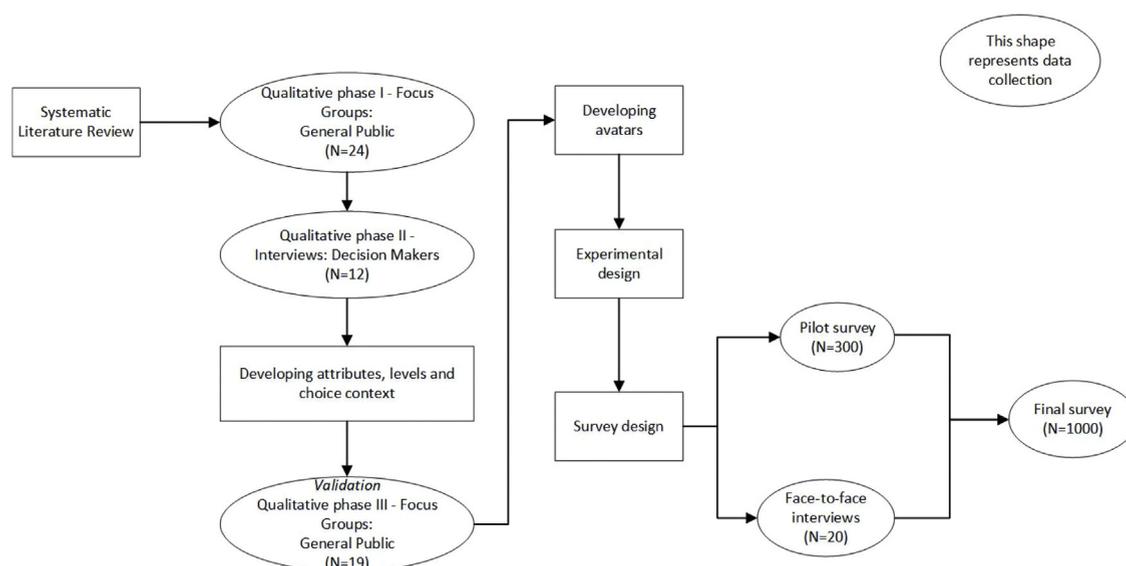
The study design is described in the next section followed by the choice modelling approach and the calculation of weights in Section 3. Results are presented and discussed in Sections 4 and 5, while Section 6 concludes.

## 2. Methods

Fig. 1 provides a schematic overview of the development of the DCE survey (depicted in boxes) and data collection (depicted in ellipses).

### 2.1. Generating attributes and informing the study design

The selection of attributes and their levels was informed by two systematic reviews of the priority setting literature and three qualitative phases of work. The first (Gu et al., 2015) synthesised what the literature has found to date regarding which attributes the general public think should count in priority setting and what weight they should



**Fig. 1.** Development of the DCE survey.

Note: Error bars represent 95% confidence intervals; life extension with reduced quality of life is used as the base level.

receive, while the second (Ghijben et al., 2017) identified factors influencing decisions made by HTA committees and assessed their importance.

Phases I and II of the qualitative work involved three focus groups ( $n=24$ ) with members of the general public (Ratcliffe et al., 2017) and 12 interviews with decision makers (drawn from current and past HTA committee members) to generate candidate attributes and levels important to priority setting in the context of HTA. Synthesizing the findings from the qualitative work with the findings from previous reviews generated five salient attributes (Table 1): age of targeted beneficiaries of treatment; two attributes in relation to severity (life expectancy without this treatment and quality of life without this treatment); and two attributes describing what would be gained with treatment (number of QALYs and the types of QALYs).

Levels were then assigned to each of these five attributes. Age levels were chosen to reflect key stages in the life cycle; we used 8 age levels, more than used in the previous literature, so as to better identify preferences across key age groups and to better reflect socio-demographic change with increased life expectancy and increasing proportions of the population in older age groups. QoL levels were presented on a 0–100% scale (e.g. Lancsar et al., 2011; Rowen et al., 2016; van de Wetering et al., 2015), ranged from very severe (5%) to mild health problems (90%) and were accompanied by a qualitative description (Table 1). Life expectancy ranged from imminent death (0–3 months) to normal life expectancy with an intentional focus on the lower end of the range to explore issues around end of life (at any age) in particular. Ranges were chosen to represent stages of life expectancy, reflecting the real world consideration of evidence on this attribute by HTA committees; ranges also characterise the reality of medical conditions better than a single number, thereby increasing external validity. The first two levels cover life expectancy up to 24 months, linking to the cri-

**Table 1**  
Attributes and Levels.

Attributes (Short name)	Levels
Age of people who will receive this treatment if treatment is funded ("age")	Infant (0–12 months)
	Child (1–12 years)
	Teen (13–17 years)
	Young adult (18–29 years)
	Adult (30–49 years)
	Older adult (50–59 years)
	Senior (60–74 years)
	Older senior (75+ years)
Quality of Life (QoL) without this treatment ("QoL")	5% (very severe health problems)
	30% (severe health problems)
	60% (moderate health problems)
Remaining life expectancy (LE) without this treatment ("LE")	90% (mild health problems)
	0–3 months
	4 months–2 years
	3–5 years
Average number of QALYs gained per person with this treatment ("QALY")	The medical condition has no effect on LE
	0.01 of a QALY
	0.5 of a QALY
	1 QALY
Type of QALYs gained with this treatment ("Type")	4 QALYs
	Life extension
	Improvement in QoL
	Mixture of life extension and improvement in QoL
	Life extension but with reduced QoL

teria used by NICE in the UK in relation to life extending end of life treatments (Chalkidou, 2012). QALY gains (the averages of QALY gain per person receiving the treatment) were informed by the range generally considered by HTA committees (drawing on interviews with decision makers in Phase II of the qualitative work) and the literature, and ranged from 0.01 of a QALY to 4 QALYs. The median QALY gain in a review of cost utility studies published in 2010 was 0.06 with 0.01 and 0.32 representing the lowest 25th and 75th percentile respectively (Wisløff et al., 2014). While treatments generating four QALYs are rarely consid-

ered by HTA committees, they are conceptually feasible. Background information presented by an avatar (discussed below) explained what the QALY levels mean in terms of extra days/weeks/years in full health. The QALY type levels included each of the two dimensions that make up a QALY in isolation; life-extending- and QoL-enhancing-QALYs. The remaining two levels combined these, one including an improvement in both dimensions and the other an extension to life combined with a reduction in QoL. The latter outcome may for example be associated with some oncology drugs and links to literature demonstrating that gains and losses in QoL are treated differently (Attema et al., 2016). A fifth level involving an increase in QoL with reduced length of life was not included as this was considered less useful to decision makers based on our qualitative work.

The inclusion of the average number of QALYs gained per person as an attribute is a novel aspect of this study. Past DCE work has presented components of a QALY gain (life years gained and/or improvements in QoL) separately and some (e.g. Lancsar et al., 2011; Rowen et al., 2016; van de Wetering et al., 2015) have combined these at the modelling stage to create a QALY variable to be used in the model estimation, but none have previously described QALYs to respondents and used this as an attribute in the experimental design. We note that Bleichrodt et al. (2005) had success describing QALYs face-to-face to participants in an economic laboratory experiment. Advantages of exploring preferences for QALYs as a whole, rather than for gains in the components of a QALY (quality and length of life) as done in earlier literature (e.g. Lancsar et al., 2011; van de Wetering et al., 2015, 2016; Shah, 2015), include that it significantly improved the statistical properties of the experimental design as it reduced the number of implausible scenarios compared to those found in earlier work (Lancsar et al., 2011). Respondents also provide preferences and values over the outcome directly of interest without needing, as required in past work, to generate a QALY at the modelling stage from disaggregated QoL and length of life attributes and to assume that the value of that post hoc generated QALY would be the same as if QALYs had been presented to members of the general public directly. Including the average number of QALYs gained as an integrated attribute also more accurately reflects the variables used in actual HTA decision making (thus enhancing external validity (Lancsar and Swait, 2014)) and allows elicitation of preferences for this key variable holistically.

Similarly, the inclusion of QALY type as an attribute is a novel addition to the literature. While the component parts of gains in QoL and length of life have been explored in earlier DCE literature (e.g. Lancsar et al., 2011; van de Wetering et al., 2015, 2016; Shah, 2015) type of QALY gain and associated weights has not been modelled. Our inclusion of QALY types as an attribute facilitates our derivation of relative weights for QALY types that can be directly applied to QALYs by decision makers in the HTA process, allows a more nuanced investigation of preferences over health gain and a head-to-head comparison of the relative value of different QALY types and their comparison to other attributes.

The focus of Phase III of the qualitative work, three focus groups (n = 19) drawn from members of the general public, was primarily to test the description and public understanding of QALY gains and different types of QALY gain, along with further refinement of the description of the choice task and attributes/levels more generally.

Given that members of the general public are unlikely to have experience with HTA decision making or QALYs, ensuring respondents' understanding of the background information provided was particularly important in this study. Considerable resources were devoted to this endeavour, including the development of an avatar-narrated background to the online survey. The avatar explained the purpose of the survey, guided respondents through the process and explained each of the attributes. QALYs and types of QALYs in particular were explained in a number of steps using animated diagrams.<sup>1</sup> Respondents' understanding was explored in the qualitative, pilot, and full study data collection as reported below.

An example choice set is provided in Appendix 1 in Supplementary material. Each choice set contained two treatments for people with medical conditions and respondents were asked: given only one treatment can be funded, which treatment would you choose to fund? Treatments labelled as A and B avoided possible framing effects linked to specific diseases. Prior to completing the task, the choice context was explained to respondents by the Avatar. This included discussion of scarcity of resources and the need for Government to make decisions about what to fund and what not to fund. A citizen framing (Dolan et al., 2003) – “so it is available to you and all Australians” – was used, thus taking an ex ante public insurance perspective. Respondents were asked to assume that any characteristics of the beneficiaries and treatment not described in the choice set were identical across the two options. Opportunity cost was explained to respondents, in that choosing to fund treatment A meant health outcomes from B would be forgone and vice versa. During consideration of each choice set respondents could refer back to the explanations and definitions provided for all attributes and levels. The online survey was further tested and refined via a pilot study (n = 300). An additional n = 20 respondents participated in face-to-face interviews following completion of the pilot survey to gain a more detailed understanding of respondents' comprehension and engagement with the survey and the tasks involved. Questions regarding respondents' understanding of QALYs in particular were included in the pilot and full study; results are reported in Section 4.

The experimental design to generate the choice data and the analysis of such data are intimately linked (Lancsar et al., 2017). Key considerations when generating the experimental design were to ensure identification of the functional forms of interest while avoiding implausible combinations of attribute levels. Best practice guidance (Johnson et al., 2013; Lancsar and Louviere, 2008) recommends the inclusion of attribute-attribute interactions

<sup>1</sup> See the avatar presented explanation of the QALY concept <https://vimeo.com/125426264> and different QALY types <https://vimeo.com/123933642>.

to avoid biased results but historically this is rarely done (recent exceptions include (Norman et al., 2013; Shah, 2015; van de Wetering et al., 2016, 2015)). Based on our systematic review of the literature (Gu et al., 2015) and our qualitative and pilot work, along with constraints imposed in the design, we included two two-way interactions, viz. interactions between age and life expectancy without treatment and between QoL without treatment and QALY type. Our a priori expectations were that the importance of life expectancy without treatment (one form of severity) would depend on the age of the potential recipients, and that preferences for QALY type would depend on severity without treatment as measured by QoL without treatment.

Choice sets were generated within Ngene 1.1.2 (ChoiceMetrics, 2014) using fixed prior coefficient values obtained from the pilot study and applying constraints (e.g. if the life expectancy without treatment is 'normal life expectancy left', then the type of QALYs gained with this treatment must be 'improvement in quality of life') to avoid implausible combinations of attribute levels.

The efficient design was generated to simultaneously accommodate two models using a model averaging approach (ChoiceMetrics, 2014). The first contained only main effects, all dummy coded except for the QALY gain attribute which was log-transformed. The second model contained main effects plus interactions between age and life expectancy (LE) and between QoL and QALY type (age and QoL entered as continuous variables). Weights 1:2 were applied to models 1 and 2 (higher weight on the model allowing for interactions) with the weighted average optimised using the D-efficiency criterion. This allowed exploration of the data in disaggregate form (model 1) to test and explore attribute functional form while also allowing estimation of a model (model 2) with the functional form of a priori interest informed by the pilot and based on our conceptual framework (discussed in Section 3.1). The final design consisted of 80 choice sets, blocked into 5 versions, each containing 16 choice sets. Respondents were randomly allocated to version (blocks).

The survey was administered via an online panel (Toluna). The panel company pays members in points which can be redeemed as vouchers or gift cards. Points can also be exchanged for prize draw tickets for cash and products. A relatively large total target sample size of 1000 respondents was set. Simulations were undertaken to inform the choice of sample size, ensuring it was sufficient to allow estimation of the models of interest accounting for respondent heterogeneity.<sup>2</sup> Quota sampling ensured that the sample was representative of the Australian general public in age and gender not only at the total sample level but also for each design block. Members of the online panel accessed an invitational web link, and after watching and listening to the avatar-narrated background, completed 16 choice

tasks plus follow-up questions including questions on the clarity of the task and the explanation of the QALY and QALY type attributes and a set of socio-demographic questions. The study was approved by the Monash University Human Research Ethics Committee.

### 3. Analysis

We took a two staged analytical approach: we estimated discrete choice models to explore preferences for each attribute level; we then used the estimated preference parameters from the preferred choice model in the calculation of relative weights for QALYs.

#### 3.1. Conceptual framework and choice models

Utility was modelled as a function of the age, life expectancy and quality of life of potential beneficiaries without treatment and the number and type of QALYs gained with treatment. The first three variables allow exploration of the relative priority of QALYs across characteristics of the beneficiary. The latter two allow exploration of the value of QALY types.

One requirement of our conceptual model was that expected utility should approach zero as QALYs approach zero; without this requirement positive utility associated with funding treatment could be attached to the other attributes such as age, for example, even in the absence of health gain (Lancsar et al., 2011). This suggests that a standard additive model for the deterministic part of utility (V) of the type often used with DCE data is not appropriate as it would mean that utility could be positive even if QALY gains were zero. Instead, the following multiplicative model accommodates this requirement:

$$V = \exp(b_1age + b_2QoL + b_3LE + b_4Type) \times QALY^{b_5} \quad (1)$$

For computational convenience, a monotonic transformation was applied to this utility function (following the approach of Lancsar et al. (2011)) to produce the following log linear model:

$$\log V = b_1age + b_2QoL + b_3LE + b_4Type + b_5\log QALY \quad (2)$$

QALY type and LE without treatment were dummy coded. The age and QoL attributes were modelled as continuous polynomials (using the midpoint of each age level) and QALY was log-transformed. The optimal functional form for these attributes was tested and determined as a quadratic function and log respectively through a series of model comparisons. This also matched the implied preference patterns from the plot of the estimated coefficients when age, QoL and QALYs were dummy coded and was consistent with the multiplicative model described above.

For notational brevity interactions were excluded from the above notation and levels suppressed. We can rewrite (2) to include the interactions between age and life expectancy and QoL and QALY type. But rather than include the main effects and interactions separately as is traditionally done, we combined the main effects of age and QoL with the interaction terms as follows

<sup>2</sup> Monte Carlo simulations (200 replications) using different sample sizes (from 500 to 1000) were used to investigate how well the generated data sets could recover the true parameter values. Simulation results suggested a sample size of 1000 recovered the true parameter values precisely (based on both MSE and MAE), including small interaction coefficients.

$$\begin{aligned}
V^* = \log(V) = & \beta_1 LE_1 + \beta_2 LE_2 + \beta_3 LE_4 + \beta_4 Type_1 \\
& + \beta_5 Type_2 + \beta_6 Type_4 + \beta_7 age \times LE_1 + \beta_8 age \times LE_2 \\
& + \beta_9 age \times LE_3 + \beta_{10} age \times LE_4 + \beta_{11} age^2 \times LE_1 \\
& + \beta_{12} age^2 \times LE_2 + \beta_{13} age^2 \times LE_3 + \beta_{14} age^2 \times LE_4 \\
& + \beta_{15} QoL \times Type_1 + \beta_{16} QoL \times Type_2 + \beta_{17} QoL \\
& \times Type_3 + \beta_{18} QoL \times Type_4 + \beta_{19} QoL^2 \times Type_1 \\
& + \beta_{20} QoL^2 \times Type_2 + \beta_{21} QoL^2 \times Type_3 \\
& + \beta_{22} QoL^2 \times Type_4 + \beta_{23} \log(QALY)
\end{aligned} \quad (3)$$

It is straightforward to show that this specification is equivalent to the traditional specification of including main effects and interactions separately (since  $LE_3 = 1 - LE_1 - LE_2 - LE_4$  and  $Type_3 = 1 - Type_1 - Type_2 - Type_4$ ),<sup>3</sup> but combining the main effects and interactions as in Eq. (3) is preferred due to its ease of interpretation as it directly specifies four types of age preferences (corresponding to four LE levels) and QoL preferences (corresponding to four QALY types), and vice versa. In particular, the preference for age when  $LE_1 = 1$  is directly estimated as  $b_7 \times age + b_{11} \times age^2$ . Likewise, the preference for age when  $LE_2 = 1$  is determined by  $b_8 \times age + \beta_{15} \times age^2$ . The preference for age when  $LE_3 = 1$  and when  $LE_4 = 1$  are determined analogously; as are preferences for QoL across the QALY type levels.

If health maximisation were important to the exclusion of all else then only the QALY attribute would be significant in the estimation of Eq. (3). Significance of age or either severity variables suggests distributional considerations are also important. Significance of QALY type variables indicates the social value of a QALY depends on how the QALY is generated. The importance of this preference, or the relative trade off between the pursuit of health maximisation versus other considerations, is explored by the relative weights.

### 3.2. Estimation

Choice data were estimated in the framework provided by random utility theory where the utility that respondent  $i$  derives from choosing alternative  $j$  in choice set  $t$  is given by

$$\begin{aligned}
U_{ijt} = & V_{ijt} + \varepsilon_{ijt} = X_{ijt}\beta + \varepsilon_{ijt}; \quad i = 1, \dots, 1000; j = 1, 2; \\
& t = 1, \dots, 16;
\end{aligned} \quad (4)$$

where  $V_{ijt}$  is the deterministic part of utility,  $\varepsilon_{ijt}$  is the random part of utility,  $X_{ijt}$  is a vector of variables representing attributes of alternative  $j$  (and some of their interactions for model 2) and  $\beta$  is a vector of coefficients. Assuming the errors  $\varepsilon_{ijt}$  are independently and identically distributed as type 1 extreme value leads to the conditional

logit model which was initially used to estimate Eq. (3). We subsequently estimated (3) with mixed logit (MIXL) (McFadden and Train, 2000) to relax the well-known restrictive assumptions of the conditional logit model; in particular to better account for the panel nature of the data while allowing for unobserved preference heterogeneity.<sup>4</sup> Under the MIXL, the utility function is given by

$$U_{ijt} = X_{ijt}\beta_i + \varepsilon_{ijt}; \quad (5)$$

where  $X_{ijt}$  is a vector of variables whose coefficients are random and specified as  $\beta_i$  representing heterogeneous preferences. Random coefficients were assumed to be uncorrelated and normally distributed except for the coefficient of the QALY attribute which was assumed to be distributed log-normally. All models were estimated in Stata 14. MIXL models were estimated by simulated maximum likelihood using the STATA command developed by Hole (2007). That command allows up to 20 random coefficients to be specified if using the automatically generated random draws. Since our preferred model specification had more than 20 random coefficients we generated our own 1000 Scrambled Halton draws (shown to outperform standard Halton draws when the dimensionality of integration is high (Bhat, 2003)) to simulate the likelihood.<sup>5</sup>

### 3.3. Weights

We calculated relative weights for each characteristic (age, QoL, LE and QALY type) as well as weights for combinations of characteristics. The former allow exploration of relative importance of levels within an attribute (e.g. different QALY types) while the latter allow exploration of the relative importance across attributes (e.g. relative importance of QALY types compared to other attributes).

We provide a conceptual overview of the method developed to calculate such weights before explaining key steps in detail. We first defined a reference case by setting the levels of each attribute approximately at their midpoints (one of the middle levels) to create an ‘‘average scenario’’: Age: 39.5 (midpoint of adult: 30–49 years); QoL: 0.6 (moderate health problems); LE: 3–5 years left; QALY: 1; Type: mixture of life extension and QoL improvement. We set the weight for this reference case equal to one. We then calculate the relative weights for each individual characteristic (or combinations of characteristics) relative to this reference case (we refer to these as ‘‘weights at the midpoints’’). Think of the reference case as the peg to which all other weights are compared. A relevant question is what impact the choice of this particular reference case has on the resulting weights. We addressed this in sensitivity analyses.

We present a new approach to the weights calculation of which, like (Lancsar et al., 2011), the compensating variation is at the core. We extend previous work in a number of ways: by allowing and accounting for respondent

<sup>3</sup>  $LE_1, LE_2, LE_3, LE_4$  refer to 0–3 months; 3 months–2 years; 3–5 years; and normal life expectancy respectively;  $Type_1, Type_2, Type_3, Type_4$  refer to QALYs that: extend life; improve QoL; are a mixture of life extension and improvement in QoL; extend life but with reduce QoL.

<sup>4</sup> A number of alternative models were also estimated but mixed logit was preferred due to interpretation, consistency with our multiplicative model and performed well on goodness of fit criteria.

<sup>5</sup> The stability of results was tested by estimating the model using 500, 600, 700, 800, 900 and 1000 draws. Stability was confirmed based on log likelihood estimates, the mean estimates and standard deviations.

heterogeneity (in the choice models, the results of which are the input to the weight calculation); by considering interactions; incorporating uncertainty into the weights calculation; by testing the impact of the reference case chosen in the weights calculation; and by undertaking formal hypothesis testing of differences between weights.

Our approach involves three parts: (1) calculation of the Hicksian compensating variation (CV) for a move from the reference case to an alternative case (where the alternative case could entail a change in a single attribute level or a change across all levels); (2) re-scaling to ensure non-negative values; and (3) normalisation of the weights such that the reference case is set to have a weight of 1. Each step is outlined in turn.

Traditionally the CV provides a measure in monetary terms of the change in welfare brought about from a change in a product/program, providing a cardinal measure of strength of preference. Here we harness the CV to value the change in a single characteristic (age etc.) or combinations of characteristics, not in monetary terms but in terms of QALYs, which also provides a cardinal measure of strength of preference. The CV is calculated as

$$CV = \frac{1}{\lambda} \left[ \ln \sum_{j=1}^J e^{V_j^1} - \ln \sum_{j=1}^J e^{V_j^0} \right] \quad (6)$$

where  $V_j^0$  and  $V_j^1$  represent the utility for each choice option  $j$  before and after the change of interest, respectively;  $J$  is the number of options in the choice set; and  $\lambda$  usually represents the marginal utility of income, but here represents the marginal utility of a QALY. The CV is interpreted as the number of QALYs that equates expected utility before and after the change of interest (denoted by the log sum values in (6)); or, put differently, values (in QALYs) the change in expected utility arising from a move from the reference case to an alternative case in which a single attribute has changed or multiple attributes have changed. Since the resulting CV values may be negative we find the smallest CV and rescale all weights to be non-negative by calculating

$$W = CV - \min(CV) \quad (7)$$

where  $\min(CV)$ , is determined by the case with approximately zero QALY gain. We then normalise all weights such that the weight for the reference case is equal to 1 by dividing through by the value of CV at the reference case to produce the final set of weights

$$W = \frac{CV - \min(CV)}{[CV - \min(CV)]_{ref}} \quad (8)$$

Noting that the CV for 'a move from' the reference case to the reference case (i.e. no change) equals 0

$$W = \frac{CV - \min(CV)}{-\min(CV)} \quad (9)$$

Due to the log functional form of the QALY variable in the utility function estimated in the discrete choice model, the marginal utility of a QALY ( $\lambda$  in Eq. (6)) is not constant and depends on the magnitude of the QALY at which it is evaluated. We calculate the marginal utility setting the

QALY level to 1, one of the middle levels. Even though the marginal utility of a QALY will differ over different QALY values which will impact the CV in (6), the relative weight will remain unchanged because the marginal utility of a QALY cancels out in the weights calculation since it enters both the numerator and denominator. As such the weights are invariant to the choice of QALY base, a contribution of this approach over earlier literature.

To investigate the impact of the choice of reference case used in the calculation of the "weights at the midpoints" we calculated a second type of weight, "weights averaging over all reference cases". In the latter all feasible reference cases are generated as all plausible combinations of the attribute levels. To take age weights for example, we allowed the attributes QoL, LE, QALY and QALY type to vary across all their levels while we fix age at "adult" as the reference whose weight will be 1. This leads to  $4 \times 4 \times 4 \times 4$  or 256 different reference cases in total. After applying the constraints, implausible cases are excluded which leads to  $k$  plausible reference cases. For each of the  $k$  cases, we find the mean CV and mean weight.<sup>6</sup> We then calculate the mean and confidence interval of these  $k$  sets of CVs and weights.

In the calculation of the weights from the chosen MIXL model we account for parameter uncertainty by using all information in the parameter distribution including the covariance matrix rather than just their mean and standard deviation. As [Hensher and Greene \(2003\)](#) note, this is more complicated because it involves simulation using the Cholesky decomposition of the covariance matrix but is preferred since using just the mean and standard deviation ignores the sampling variance in the point estimates. This approach allowed us to obtain a sample of draws from the asymptotic distribution of the mean CV and weight from which we use the percentiles (standard deviation) as its confidence interval (standard error). Finally, hypothesis testing was undertaken to test for statistically significant differences across weights using a two-tailed test.

#### 4. Results

The sample of 1000 respondents is representative of the Australian population by age and gender. Further comparison to the population norm for other characteristics is shown in Appendix 2 in Supplementary material, suggesting the sample and population are also similar in education, household income, employment status, marital status and place of residence. We do not have information on the sample's political views or views on priority setting.

We explored respondents' understanding of and engagement with the task in various ways in the full study and pilot. In the full study, over 88 % found the task to be clear or very clear, 9 % neither clear/unclear and only 2.7 % found it unclear. Very similar results (89 %) were found in the pilot. In the pilot, 95.3 % of respondents felt the explana-

<sup>6</sup> Note that the number of plausible reference cases differ by attributes: age: 179 plausible references (out of 256 scenarios); QoL: 368 plausible reference cases (out of 512 scenarios); LE: 472 plausible reference cases (out of 512 scenarios); QALY type: 384 plausible references (out of 512 scenarios).

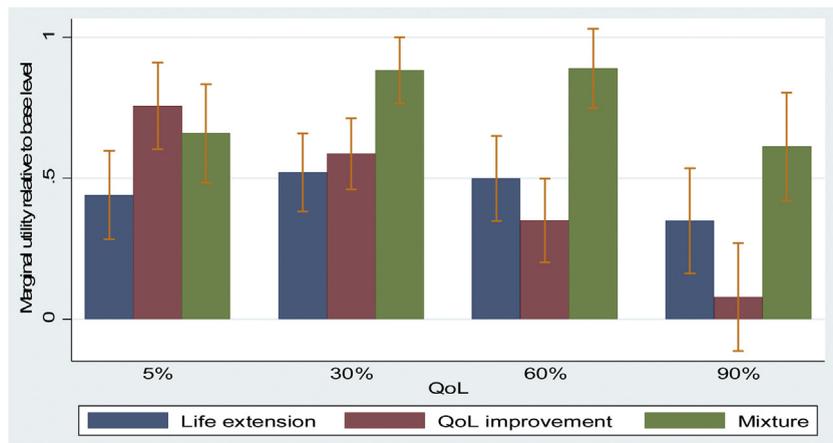


Fig. 2. Relative importance of QALY types accounting for the interaction with QoL without treatment, holding QALY gain constant.

tion of QALYs and QALY types provided in diagrams and by the avatar was clear. When asked in the pilot interviews if they would prefer to have the avatar or just read the information, 17 of the 20 interviewees (85 %) said they would prefer to have the avatar.

#### 4.1. Choice models

The preferred model in terms of consistency with the proposed multiplicative model, interpretation and performed well in terms of goodness of fit was the MIXL estimation of Eq. (3). Parameter results for that model and for the base logit model are provided in Appendix 3 in Supplementary material.<sup>7</sup>

All attributes are statistically significant predictors of choice. There is significant unobserved heterogeneity as indicated by the significant standard deviations. Larger QALY gains are preferred but at a diminishing rate. This is not surprising given QALYs entered the model in log form. However, the same result holds when QALY levels were included as dummies in either the full study data or the pilot data. To help interpret the main effects and interaction results presented in Appendix 3 in Supplementary material and to elucidate the relationship between the attribute levels and choice, we harness the MIXL results to generate a number of plots. All figures plot marginal utility accounting for both main effects and interactions and hold the QALY gain constant. The first (Fig. 2) explores preferences for the type of QALY gain accounting for its interaction with QoL without treatment. Across all four QoL levels, QALYs that extend life but reduce QoL (the base level) are always the least preferred; the exception being when QoL without treatment is 90 %, in which case

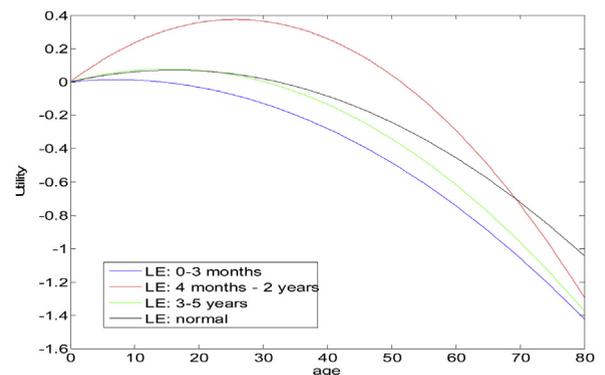


Fig. 3. Relative importance of age groups accounting for the interaction with life expectancy.

QALYs that extend life but reduce QoL are not statistically significantly different from QALYs that improve QoL only. QALYs generated as a mixture of life extension with QoL improvement are always preferred over QALYs that only improve QoL with no change in LE except when QoL without treatment is 5 % in which case the confidence intervals largely overlap. QALYs generated as a mixture of life extension and QoL improvement are also preferred over QALYs generated solely by life extension in general except for when QoL without treatment is 5 % or 90 % in which case the confidence intervals overlap. QoL-improving QALYs are preferred over life-extending QALYs when QoL is low (5 %), while for other QoL levels the confidence intervals overlap meaning we cannot rule out that they are not statistically significantly different.

Fig. 3 explores the relative importance of different age groups accounting for the significant interaction between age of recipient and life expectancy without treatment; depicted in four plots of the functional form of age (at the mean), one for each of the four life expectancy levels. For all four LE levels, the plots are concave meaning that the utility from choosing to fund a treatment increases as age of recipient increases until a maximum (most preferred age) after which it decreases as age increases. The most preferred age group (defined in Table 1) depends on remaining

<sup>7</sup> We also undertook analysis including a random alternative specific constant in the model which was not statistically significant indicating an absence of a significant preference for left or right position of the alternatives in the choice set (significant standard deviation suggests heterogeneity around the mean result). We also tested for scale differences across choice tasks 1-4, 5-8, 9-12 and 13-16 by estimating a heteroscedastic conditional logit and found that position of the choice tasks did not impact scale.

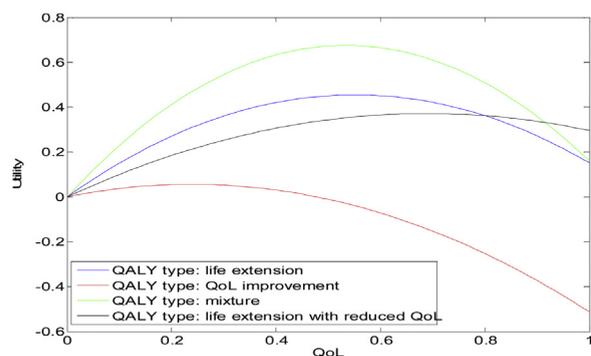


Fig. 4. Relative importance of QoL accounting for the interaction with QALY type.

life expectancy without treatment. When life expectancy is extremely short (0–3 months) infants and children are prioritised, when it is relatively short (4 months–2 years) young adults are prioritised but when LE is longer (3–5 years or normal) teens are prioritised.

Fig. 4 depicts the relative importance across QoL without treatment levels accounting for the significant interaction with QALY type. The figure includes four plots of the functional form of QoL without treatment, one for each QALY type. The utility of choosing to fund a treatment generally increases as QoL without treatment increases until a maximum after which it decreases as QoL increases. Interestingly, this means the most severe conditions are not uniformly prioritised. The most preferred level of QoL differs depending on the type of QALY that would be gained with treatment. In particular, when the treatment generates QALYs made up of QoL improvement only, patients who have very low QoL without treatment (i.e., the most severe) are prioritised. When the treatment extends life or both extends life and improves QoL, patients in moderate QoL without treatment are prioritised. When treatment extends life but simultaneously reduces QoL, patients in relatively high QoL without treatment are prioritised.

## 4.2. Weights

### 4.2.1. Weights for individual characteristics

The distributional weights at the midpoints (specifically at one of the middle levels on each of the attributes) (columns 3–5) and averaging over all bases (columns 6–8) are presented in Table 2 which accounts for main effects and interactions. Overall the size of weights within an attribute were similar across the two sets of weights. Taking into account statistical significance, the ordering between attribute levels from the two sets of weights are nearly identical, suggesting choice of midpoints as the reference case was reasonable.

Within each attribute, the reference level has a weight of one. A weight greater (less) than one indicates that that level is valued more (less) highly than the reference case. The magnitude of the weight relative to 1 indicates the strength of that preference. Within Table 2 if the confidence intervals do not overlap the value of 1, the estimated weights are statistically different from 1. Where they do

overlap, we formally tested the statistical difference using two-tailed hypothesis tests.

Focusing on QALY types, QALYs generated by a mixture of life extension and QoL improvement are valued more than those generated by either of these components in isolation; and both of these are weighted more highly than those QALYs which extend life but at the expense of reduced QoL. All QALY type weights are statistically significantly different from each other except for the weights for QALYs generated from QoL in isolation and LE in isolation. A QALY that extends life but reduces QoL has a weight of approximately half (0.53) that of a QALY generated via a mixture.

For age, the smallest weights are for gains for older age groups, and progressively so. For example, on average a gain for an infant is weighted almost 2.5 times that of an older senior. Weights for infant, child, teen and young adult are not statistically significantly different to each other, but, based on hypothesis testing, are all statistically significantly greater than the weights for adult, older adult, senior and older senior. The weight for adults is statistically significantly different from the weights for all older groups. All remaining weights are statistically significantly different to each other. On average, an adult is weighted more than twice that of an older senior.

For QoL without treatment, all weights are statistically significantly different to 1 (the reference of 60 %) except QoL without treatment of 90 %. All weights are statistically significantly different from each other except weights for QoL of 30 % and 90 %. The largest significant weight (with a value of 1) is given to those who are in reasonable QoL (60 %) and lowest, a weight of 0.64, to those in poorest QoL = 5 % (severe health problems).

Looking at LE without treatment, again those in the most severe conditions are not prioritised, including those at the end of their life (0–3 months to live). However, none of the LE weights are statistically significantly different to the reference case of 3–5 years. But, the weight for 0–3 months remaining LE (0.92) is statistically significantly less than the weight for 4 months to two years (1.08) which is significantly greater than the weight for normal life expectancy (0.90).

### 4.2.2. Weights for combinations of characteristics

To explore the relative importance across the attributes we estimated mean weights for all plausible combinations of selected attribute levels and present these in Appendix 4 in Supplementary material. This used four levels for age (infant, teen, adult, older senior), three levels for QoL (5 %, 60 %, 90 %), three levels for life expectancy (0–3 months, 3–5 years, normal LE), three QALY type levels (life extension, QoL improvement, mixture) and a QALY gain of 1. This produced 108 possible combinations in total, of which 68 are plausible.

General findings include that age levels of teen and infant, the QALY type 'mixture', QoL of 60 % and 90 %, and life expectancy of 3–5 years, in general lead to larger weights. Focusing on scenarios which involve a QALY gain of 1, the largest weight (of 1.13) was for a scenario involving a teen in moderate health (60 %) with 3–5 years to live who receives 1 QALY that is made up of a mixture of life exten-

**Table 2**

Mean characteristics weights at the midpoints and averaged over all bases (accounting for main effects and interactions).

	At mid points			Averaging over all bases		
	mean	95 % CI		mean	95 % CI	
<i>Age</i>						
Infant	1.0725	<b>0.9687</b>	<b>1.1880</b>	1.0014	<b>0.9274</b>	<b>1.0741</b>
Child	1.1065	1.0253	1.2057	1.0559	1.0021	1.1191
Teen	1.1262	1.0682	1.2074	1.1008	1.0572	1.1602
Young Adult	1.1122	1.0695	1.1694	1.1041	1.0685	1.1537
Adult = reference	1	.	.	1	.	.
Older Adult	0.8158	0.7277	0.8791	0.8010	0.7031	0.8652
Senior	0.6331	0.4542	0.7621	0.6023	0.4050	0.7333
Older Senior	0.4452	0.1650	0.6441	0.4064	0.1220	0.6053
<i>QoL without treatment</i>						
5 %	0.6392	0.4216	0.7917	0.7359	0.5921	0.8311
30 %	0.8954	0.8170	0.9517	0.9189	0.8672	0.9529
60 % = reference	1	.	.	1	.	.
90 %	0.8881	<b>0.7468</b>	<b>1.0003</b>	0.9520	<b>0.8839</b>	<b>1.0155</b>
<i>LE without treatment</i>						
0 to 3 months	0.9183	<b>0.8062</b>	<b>1.0136</b>	0.9760	<b>0.8895</b>	<b>1.0834</b>
4 months to 2 years	1.0778	<b>0.9888</b>	<b>1.1831</b>	0.9483	<b>0.8733</b>	<b>1.0061</b>
3–5 years = reference	1	.	.	1	.	.
normal LE	0.8989	<b>0.7583</b>	<b>1.0289</b>	0.9368	<b>0.7935</b>	<b>1.0759</b>
<i>QALY Type</i>						
Life extension	0.7570	0.5904	0.8768	0.7848	0.6629	0.8727
QoL improvement	0.6842	0.4982	0.8213	0.7938	0.6601	0.8910
Mixture = reference	1	.	.	1	.	.
Life extension with reduced QoL	0.5321	0.2981	0.6991	0.5566	0.3386	0.6974

Note: Confidence intervals in bold includes 1 suggesting the corresponding weight is not statistically different from 1.

sion and improvement in QoL. All weights that are greater than one, meaning they are valued more highly than the reference case, are for the QALY type 'mixture'. The smallest weight (of 0.19) was for a scenario involving an older senior in severe health (5 % QoL without treatment) who has 3–5 years to live without treatment and receives 1 QALY with treatment generated purely by life extension.

Within scenarios, age followed by QALYs and QALY type are the main drivers of the results, followed by the severity measures life expectancy and QoL without treatment. Replicating the scenarios but varying QALYs over values of 0.01, 0.5, 1, and 4 indicates that, as QALYs increase, the number of weights greater than 1 also increase; meaning that the alternative is preferred to the reference case. All else equal, the increment in weights from QALY=0.01 to QALY=0.5 is much larger (around three times larger or more depending on the other four attributes) than the increments from QALY=0.5 to QALY=1 and from QALY=1 to QALY=4. All else equal, the increment (marginal increase) in weights from QALY=0.5 to QALY=1 and from QALY=1 to QALY=4 are in general small. This is consistent with the shape of the log functional form of QALY gain attribute (increasing at a decreasing rate) but also generally consistent with the functional form implied by plotting the coefficients for dummies for each level.

## 5. Discussion

This study is the first to explore the relative value of different QALY types via its inclusion as an attribute in a DCE. We have made a novel contribution to the existing empirical evidence suggesting that not all QALYs are equal, by generating relative weights for different types of QALYs and their relative priority compared to other characteris-

tics across which social values for a QALY could differ. The significantly different weights across QALY types leads us to reject the hypothesis that a 'QALY is a QALY is a QALY' and indicates that the general public care about the composition of the QALYs gained with treatment. Specifically, for both the weights for individual characteristics and combinations of characteristics, the lowest weight is given to QALYs that involve life extension but at the cost of reduced QoL and the highest weight is given to QALYs generated by a mix of life extension and QoL improvement. While the general public clearly favour treatments that generate mixed QALYs, such preferences depend on the underlying severity as measured by QoL without treatment. For example, when QoL without treatment is low, QALYs generated by improvements in QoL receive highest weight. In contrast to Pennington et al. (2015) and Mason et al. (2009) who derived a higher monetary value for life extending gains rather than QoL-enhancing gains (about 75 % higher for Pennington et al. and 66 % higher for Mason et al), we found that while the weight for life extending QALYs was greater than that for QoL-enhancing QALYs, the difference was not statistically significant. In addition to the different elicitation methods used across those studies and ours, the difference in results may also relate to our exploration of the relative priority of different QALY types in the presence of other characteristics rather than in isolation.

We have also added to the empirical literature exploring the trade-off between health maximisation and equity considerations by calculating weights for characteristics of the recipients of treatment, viz. age and severity. Not surprisingly, more QALYs are preferred to less but at a diminishing rate, results which are consistent with a health maximisation approach found elsewhere (e.g. Lancsar et al., 2011; Rowen et al., 2016; Skedgel et al., 2015). However, signifi-

cant weights on other attributes suggest members of the general public are prepared to trade off health gain for other distributional considerations, meaning health gain matters, but it is not the only maximand. Consistent with much of the literature (Gu et al., 2015), the young received more weight than older age categories but the very young are not the most preferred and there was little distinction between the younger age groups. The preference for giving priority to the young was strongest when life expectancy is very short.

Severity was defined based on severity of the condition without treatment, or prognosis. Interestingly, the largest weight is not given to those with the most severe conditions, measured in terms of either QoL or life expectancy without treatment. Instead, those in moderate health were prioritised. This may link to the idea that there are thresholds in health states above which health gains become meaningful in that they can change what the recipients of the health gain are able to do, for example in terms of returning to work, reducing care burden on families etc. A similar point has been raised by others (e.g. Shah, 2009) and warrants further investigation. This finding that the most severe are not prioritised is also consistent with recent work on social preferences (e.g. Dolan and Tsuchiya, 2005; Lancsar et al., 2011; Norman et al., 2013; Skedgel et al., 2015; van de Wetering et al., 2015) but runs counter to the broader literature in which the most severe are generally prioritised (Gu et al., 2015; Nord and Johansen, 2014). It is important to note that when investigating this issue, a number of studies did not allow for multi attributes to be traded off simultaneously and most other studies have not taken interactions into account (Gu et al., 2015), with some recent exceptions noted earlier. Accounting for interactions indicated that those in most severe health states are prioritised for particular types of QALY gains and for particular age groups. Specifically, the most severe, in terms of lowest QoL without treatment, are prioritised when treatment improves QoL only. In terms of life expectancy without treatment, interestingly, given NICE's end of life premium (Chalkidou, 2012), those at the end of their life are not prioritised except if the recipients of treatment are infants, in which case the shortest LE (most severe) are prioritised. This is in contrast to results found by Rowen et al. (2016) but consistent with the general finding in the literature of a lack of preference for end-of-life premium (Gu et al., 2015; Gyrd-Hansen, 2018; Shah, 2018). In terms of the relative importance across the attributes, the weights provide a measure of strength of preference and indicate the rank ordering: age, QALY gain, QALY type, severity measures.

## 6. Conclusion

We set out to build a framework to generate robust and useful relative QALY weights consistent with calls for improved methodological approaches in this area (Gu et al., 2015; Lancsar et al., 2011). Methodologically, we developed a new approach to the weights calculation. Past work on relative social values has generally assumed homogeneous preferences (with exceptions such as van de Wetering et al. (2015)) and accounted for neither heterogeneity nor uncertainty in the calculation of the distributional weights. We

addressed both here. We explored the robustness of our results at both the stage of choice modelling and weights calculation. Importantly, we tested the impact of the choice of reference case used in the weights calculation. Comparing the weights calculated at the midpoint of the attribute levels to the weights calculated by averaging over feasible reference cases demonstrated that the choice of reference did not have a major impact on the results.

Considerable effort was expended on the 'front end' of this study, particularly via in depth qualitative work and pilot testing. Given the nature of the topic, we used multimedia approaches, including an animated avatar, to explain the background and key concepts used in the DCE. The use of QALYs and QALY types as attributes in a DCE is novel to this study and validity testing suggested they were generally understood and well received by respondents. This inclusion of QALYs (rather than its component parts) also allowed for a more robust experimental design with many fewer implausible scenarios than past work (Lancsar et al., 2011). Importantly, interactions were included in the experimental design, which, as our discussion of the empirical results highlights, allowed more nuanced investigation of preferences and weights. We also applied a more informative way of modelling interactions. As such, we have endeavoured to extend the previous literature in a number of methodological dimensions.

As is usual there are a number of limitations to this work and avenues for further research. Although the sample is representative in terms of demographics, we do not know the political views of respondents, nor their views on priority setting (Reckers-Droog et al., 2018). Future work could consider attitudes toward health care resource allocation as a sampling variable and draw on past work to do so (e.g. Mason et al., 2016). We deliberately used a range to describe the levels on the QoL without treatment attribute as it best fit out research questions, most closely aligned with how medical conditions are described and with the nature of information used in HTA decision making, improving external validity. A potential limitation is respondents interpretation of the levels may differ within such ranges. Under our QALY type attribute we included four levels including the case where QALYs are generated via an extension to life but a reduction in QoL. The reverse case, QALYs that increase QoL but reduce life expectancy, may also be of interest. Work by McNeil et al. (1981) explored QoL and survival tradeoffs and found, at the individual level, a substantial minority of patients were prepared to trade-off reduced survival to maintain their QoL in terms of not losing their voice. An interesting question is whether the general public would hold similar preferences in relation to the allocation of scarce societal healthcare resources. Future methodological work also could test in a head to head fashion if the same values are derived from the presentation of QALYs directly in the choice sets as we have done here compared to the earlier approach in the literature of presenting gains in QoL and life expectancy separately with researchers generating QALYs at the modelling stage. As noted above, we viewed choice of the reference case as sufficiently important to test the implications of our choice. Future research could identify from the general public, and/or decision makers, their

views on an appropriate reference case (Wouters et al., 2015). Choice of reference case could indeed depend on the allocation decision or cost effectiveness/utility analysis at hand, but a counter argument would be the use of a common base to allow for uniformity in decision making. Accounting for additional attribute interactions would be another natural extension. As is the standard approach, we modelled choice using a random utility maximisation model which is a reasonable choice from a societal perspective. While our qualitative work, including interviews following completion of the pilot survey with a subset of respondents, did not suggest other decision rules were being used, interesting future research could test whether other decision rules such as regret minimisation (e.g. de Bekker-Grob and Chorus, 2013) play a role in decision making in relation to the allocation of societal health care resources. More generally other settings in which to apply the methods used in this study are with decision makers themselves and in other HTA jurisdictions.

In terms of implications for policy, these results suggest that the Australian public would trade off health gain for considerations in relation to the type of gain and the age of the beneficiary but less weight (compared to age and type of QALY) is given to the pre-treatment severity of the recipient of the QALYs. There also does not appear to be an appetite for additional weighting for treatments targeted at those at end of life. We do not view our results as negating the basic premise of a QALY that length of life and quality of life can be aggregated in one measure nor that QALYs need to be discarded. Rather, we interpret our results as implying that the social value attached to that aggregate measure will differ depending on its type and beneficiary meaning fixed QALY values (or equivalently fixed cost-per-QALY thresholds) may not reflect societal preferences. This is also consistent with preferences revealed by health technology assessment bodies through their own policies (e.g. NICE's end of life criteria; PBAC's rule of rescue) which demonstrate funders are prepared to move away from a fixed value of a QALY depending on the type of QALY or beneficiary of the QALY. We provide in our paper a method for how to include such social values into HTA decision making.

#### CRedit authorship contribution statement

**Emily Lancsar:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing - original draft. **Yuanyuan Gu:** Conceptualization, Formal analysis, Methodology, Software, Writing - review & editing. **Dorte Gyrd-Hansen:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Jim Butler:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Julie Ratcliffe:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Liliana Bulfone:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing. **Cam Donaldson:** Conceptualization, Funding acquisition, Methodology, Writing - review & editing.

#### Declaration of Competing Interest

None.

#### Acknowledgements

This study was funded by an Australian National Health and Medical Research Council project grant APP1047788. The authors thank the Editor and three anonymous referees for helpful comments and suggestions. The authors thank the study advisory committee (Vic Adamowicz, Andrew Mitchell, Lloyd Sansom and Rosalie Viney) as well as Helen Mason, David Johnston, Jan Abel Olsen, Jose Luis Pinto Prades, Joffre Swait, Nada Wasi and John Wildman for useful discussions. We also thank Kate Ennis for excellent research assistance during part of the qualitative phase of the study.

#### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jhealeco.2020.102303>.

#### References

- Attema, A.E., Brouwer, W.B., l'Haridon, O., Pinto, J.L., 2016. An elicitation of utility for quality of life under prospect theory. *J. Health Econ.* 48, 121–134.
- Australian Department of Health, Canberra 2017. *Health Technology Assessment Consumer Consultative Committee Terms of Reference*.
- Baker, R., Bateman, I., Donaldson, C., Jones-Lee, M., Lancsar, E., Loomes, G., Mason, M., Odejar, M., Pinto Prades, J.-L., Robinson, A., Ryan, M., Shackley, P., Smith, R., R S, Wildman, J., 2010. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY project. *Health Technol. Assess.* 14.
- Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomised and scrambled Halton sequences. *Transp. Res. Part B Methodol.* 37, 837–855.
- Bleichrodt, H., Doctor, J., Stolk, E., 2005. A non parametric elicitation of the equity-efficiency trade-off in cost-utility analysis. *J. Health Econ.* 24, 655–678.
- Bobinac, A., van Exel, J., Rutten, F., Brouwer, W., 2012. Inquiry into the relationship between equity weights and the value of the QALY. *Value Health.* 1119–1126.
- Brazier, J., Ratcliffe, J., Saloman, J., Tsuchiya, A., 2017. *Measuring and Valuing Health Benefits for Economic Evaluations*. OXFORD University Press.
- Chalkidou, K., 2012. Evidence and values: paying for end-of-life drugs in the British NHS. *Health Econ. Policy Law* 7, 393–409.
- ChoiceMetrics, Australia 2014. *Ngene User Manual & Reference Guide*. (Ed)(Eds).
- Cookson, R., Drummond, M., Weatherly, H., 2009. Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Econ. Policy Law* 4, 231–245.
- Culyer, A., 2006. Nice's use of cost effectiveness as an exemplar of a deliberative process. *Health Econ. Policy Law* 1, 299–318.
- de Bekker-Grob, E., Chorus, C., 2013. Random regret-based discrete-choice modelling: an application to healthcare. *Pharmacoeconomics* 31, 623–634.
- Dolan, P., Tsuchiya, A., 2005. Health priorities and public preferences: the relative importance of past health experience and future health prospects. *J. Health Econ.* 24, 703–714.
- Dolan, P., Olsen, J.A., Menzel, P., Richardson, J., 2003. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ.* 12, 545–551.
- Donaldson, C., Wright, K., 1990. Programme-specific QALYs: a reply. *J. Health Econ.* 8, 489–491.
- Donaldson, C., Atkinson, A., Bond, J., Wright, K., 1988. Should QALYs be programme-specific? *J. Health Econ.* 7, 239–257.

- Ghijben, P., Gu, Y., Lancsar, E., Zavarsek, S., 2017. Revealed and stated preferences of decision makers for priority setting in health technology assessment: a systematic review. *Pharmacoeconomics*, 1–18.
- Gu, Y., Lancsar, E., Ghijben, P., Butler, J.R., Donaldson, C., 2015. Attributes and weights in health care priority setting: a systematic review of what counts and to what extent. *Soc. Sci. Med.* 146, 41–52.
- Gyrd-Hansen, D., 2018. Is there additional value attached to health gains at the end of life? A revisit. *Health Econ.* 27, e71–e75.
- Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. *Transportation* 30, 133–176.
- Hole, A.R., 2007. Estimating mixed logit models using maximum simulated likelihood. *Stata J.* 7, 388–401.
- Johnson, F.R., Lancsar, E., Marshall, D., Kilambi, V., Muhlbacher, A., Regier, D.A., Bresnahan, B.W., Kanninen, B., Bridges, J.F.P., 2013. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health* 16, 3–13.
- Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making a user's guide. *Pharmacoeconomics* 26, 661–677.
- Lancsar, E., Swait, J., 2014. Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics* 32, 951–965.
- Lancsar, E., Wildman, J., Donaldson, C., Ryan, M., Baker, R., 2011. Deriving distributional weights for QALYs through discrete choice experiments. *J. Health Econ.* 30, 466–478.
- Lancsar, E., Fiebig, D.G., Hole, A.R., 2017. Discrete choice experiments: a guide to model specification, estimation and software. *Pharmacoeconomics* 35, 697–716.
- Mason, H., Jones-Lee, M., Donaldson, C., 2009. Modelling the monetary value of a QALY: a new approach based on UK data. *Health Econ.* 18, 933–950.
- Mason, H., van Exel, J., Baker, R., Brouwer, W., Donaldson, C., 2016. From representing views to representativeness of views: illustrating a new (Q2S) approach in the context of health care priority setting in nine European countries. *Soc. Sci. Med.* 166, 205–213.
- McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *J. Appl. Econ.* 15, 447–470.
- McNeil, B., Weichselbaum, R., Pauker, S., 1981. Speech and survival: tradeoffs between quality and quantity of life in laryngeal cancer. *N. Engl. J. Med.* 305, 982–987.
- Miyamoto, J., Eraker, S., 1989. Parametric models of the utility of survival duration: tests of axioms in a generic utility framework. *Organ. Behav. Hum. Decis. Process.* 44, 166–202.
- NICE, 2008. Social Value Judgements. Principles for the Development of NICE Guidance, London.
- Nord, E., Johansen, R., 2014. Concerns for severity in priority setting in health care: a review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. *Health Policy* 116, 281–288.
- Norman, R., Hall, J., Street, D., Viney, R., 2013. Efficiency and equity: a stated preference approach. *Health Econ.* 22, 568–581.
- Pennington, M., Baker, R., Brouwer, W., Mason, H., Hansen, D.G., Robinson, A., Donaldson, C., Team, E., 2015. Comparing WTP values of different types of QALY gain elicited from the general public. *Health Econ.* 24, 280–293.
- Pliskin, J., Shepard, D., Weinstein, M., 1980. Utility functions for life years and health status. *Oper. Res.* 28, 206–224.
- Ratcliffe, J., Lancsar, E., Walker, R., Gu, Y., 2017. Understanding what matters: an exploratory study to investigate the views of the general public for priority setting criteria in health care. *Health Policy* 121, 653–662.
- Reckers-Droog, V.T., Van Exel, N.J.A., Brouwer, W.B.F., 2018. Who should receive treatment? An empirical enquiry into the relationship between societal views and preferences concerning healthcare priority setting. *PLoS One* 13 (6), <http://dx.doi.org/10.1371/journal.pone.0198761>.
- Round, J., Paulden, M., 2018. Incorporating equity in economic evaluations: a multi-attribute equity state approach. *Eur. J. Health Econ.* 19, 489–498.
- Rowen, D., Brazier, J., Mukuria, C., Keetharuth, A., Risa Hole, A., Tsuchiya, A., Whyte, S., Shackley, P., 2016. Eliciting societal preferences for weighting QALYs for burden of illness and end of life. *Med. Decis. Mak.* 36, 210–222.
- Shah, K., 2009. Severity of illness and priority setting in healthcare: a review of the literature. *Health Policy* 93, 77–84.
- Shah, K., 2015. Valuing health at the end of life: a stated preference discrete choice experiment. *Soc. Sci. Med.* 125, 48–56.
- Shah, K., 2018. Valuing health at the end of life: a review of stated preference studies in the social sciences literature. *Soc. Sci. Med.* 204, 39–50.
- Skedgel, C., Wailoo, A., Akehurst, R., 2015. Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. *Med. Decis. Mak.* 35, 94–105.
- van de Wetering, L., van Exel, J., Bobinac, A., Brouwer, W.B.F., 2015. Valuing QALYs in relation to equity considerations using a discrete choice experiment. *Pharmacoeconomics* 33, 1289–1300.
- van de Wetering, E., van Exel, J., Rose, J., Brouwer, W., 2016. Are some QALYs more equal than others. *Eur. J. Health Econ.* 17, 117–127.
- Wagstaff, A., 1991. QALYs and the equity-efficiency tradeoff. *J. Health Econ.* 10, 21–41.
- Weinstein, M.C., 1988. A QALY is a QALY is a QALY – or is it? *J. Health Econ.* 7, 289–290.
- Williams, A., 1997. Intergenerational equity: an exploration of the 'fair innings' argument. *Health Econ.* 6.
- Wisløff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., Olsen, J.A., 2014. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics* 32, 367–375.
- Wouters, S., van Exel, J., Rohde, K., Brouwer, W., 2015. Are all health gains equally important? An exploration of acceptable health as a reference point in health care priority setting. *Health Qual. Life Outcomes* 13, 79.