

## **Measuring and valuing health and quality of life using the EQ Health and Wellbeing Short: perspectives from members of the NICE Public Involvement Programme Expert Panel**

Mukuria C<sup>1</sup>, Peasgood T<sup>2</sup>, Devlin N<sup>2</sup>, Powell P<sup>1</sup>, Carlton J<sup>1</sup>, McDool E<sup>1</sup>, Brazier J<sup>1</sup>, Rowen D<sup>1</sup>

1. University of Sheffield

2. University of Melbourne

Corresponding author: C.mukuria@sheffield.ac.uk

### **Abstract**

**Objective:** The EQ Health and Wellbeing Short (EQ-HWB-S™ Experimental version) has been developed to support decision-making in health and social care. The measure has nine dimensions: mobility, activities, exhaustion, loneliness, cognition, anxiety, sadness/depression, control, and pain. Utility weights have been generated in a feasibility study using a modified EuroQol Valuation Technology protocol. Members of the public are key stakeholders who are impacted by the use of measures such as the EQ-HWB-S in decision-making. Therefore, the aim of this stage of development was to gain their views regarding the EQ-HWB-S, including the weights, for decision-making.

**Method:** Members of the NICE Public Involvement Programme Expert Panel were invited to participate, with volunteers selected to represent varying age, gender, health and caring responsibilities. To familiarise the group with the measure and the source of the weights, each person completed a valuation interview (time trade-off (TTO) and discrete choice experiment (DCE)). This was followed by a cognitive debrief and information giving group session, where the weights from the feasibility valuation study were presented. Two subsequent separate focus groups obtained views regarding the measure, the utility weights, the sample (including exclusions) and the methods used. All sessions took place online. Focus groups were recorded, transcribed, and analysed using a framework approach.

**Results:** Twelve people (50% female, aged 28-74) completed the interviews and nine attended the focus groups. EQ-HWB-S was viewed positively due to the inclusion of dimensions such as exhaustion and loneliness. Some missing dimensions were identified (e.g. coping, sleep), but existing dimensions were considered to cover some of these (e.g. sleep covered by exhaustion). There was surprise at the small utility decrements for anxiety, control and exhaustion relative to other dimensions. Weights were seen as reflecting societal norms, respondent experience or knowledge, the composition of the sample and the interpretation of items. There were concerns that the valuation survey sample was not diverse or large enough to adequately represent the values of those who would be impacted by decisions based on EQ-HWB-S. Participants only supported data exclusions where it could be evidenced by multiple sources that the respondent did not understand or fully engage in the exercise. Other exclusions were considered problematic either because the data could reflect true preferences or for ethical reasons. DCE

was preferred to TTO, but participants suggested TTO could be improved by providing more background information, different practice states, particularly the wheelchair state, and offering post-survey debriefs.

Conclusion: The EQ-HWB-S was viewed positively by informed members of the public, but there were concerns regarding the utilities and their source. This innovative study provides insight from important stakeholders whose views are rarely sought once valuation studies are completed.

**Funding:** This study is independent research funded by the UK Medical Research Council (Grant number 170620), and the EuroQol Research Foundation.

**Disclaimer:** The views expressed in this manuscript are those of the authors and not necessarily of our funders, the National Institute of Health and Care Excellence (NICE), the Department of Health and Social Care or those acknowledged.

**Acknowledgement:** We would like to thank the NICE Public Involvement Group Expert Panel members who took part in this study, NICE for facilitating recruitment and the National Institute for Health Research Yorkshire and Humber Applied Research Collaboration for funding support.

## **BACKGROUND**

A new measure of health and quality of life, the EQ Health and Wellbeing (EQ-HWB™), has been developed to support decision-making in health and social care, including for informal carers[1]. There are two versions, a profile measure with 25 questions and a shorter version with 9 questions, the EQ-HWB Short (EQ-HWB-S™). The shorter EQ-HWB-S was developed so that it could be scored using valuation techniques on a utility scale anchored at 0 (dead) to 1 (full health and quality of life). Techniques to generate utility values can be cognitively burdensome, therefore shorter measures are needed to ensure that participants can engage meaningfully with the task and provide valid values. Utilities are combined with length of life to generate quality adjusted life years (QALYs). QALYs are used in economic evaluation to assess changes in health and quality of life and/or length of life and are therefore a useful metric for assessing the impact of different interventions. QALYs are the recommended metric for use by reimbursement agencies such as the National Institute of Health and Care Excellence (NICE)[2]. However, health measures may be limited in their assessment of some patient populations[3] and they may not always cover aspects that are important in other contexts such as in social care where health may remain the same or deteriorate, but interventions may improve aspects such as independence which go beyond health[4]. The EQ-HWB measures were developed to address some of these limitations.

The development of the EQ-HWB measures drew upon current best practice [5, 6]. Potential dimensions were identified based on the views of service users (patients, informal carers and social care users)[7] and a potential pool of questions was identified and tested qualitatively [8] and quantitatively[5]. The final stage was a preference-elicitation or valuation study to generate utility values. A mixed methods pilot was undertaken to test whether the EQ-HWB-S could be valued using standard valuation techniques based on a modified version of the EuroQol Valuation Technology (EQ-VT v2) protocol developed to value the EQ-5D-5L[9]. The pilot assessed engagement for the EQ-HWB-S compared to the EQ-5D-5L when using composite time trade-off (cTTO) and alternative presentations of paired choices for discrete choice experiments (DCE). The mixed method pilot indicated that the EQ-HWB-S could be valued using this approach. A feasibility study (n=521) was undertaken to value the EQ-HWB-S. This yielded an initial value set that can be used to generate utility values for the EQ-HWB-S.

At every stage, different stakeholders were involved in the development of the measure including members of the public, patients and informal carers (referred to as patient and public

involvement and engagement (PPIE)) [10]. The value of PPIE within the development and refinement of outcome measures (including preference-based measures) is becoming increasingly recognized[11, 12]. PPIE are partners (rather than research participants) and they can be involved in all or some aspects of research from developing research proposals to interpreting data and writing up. The project PPIE group provided valuable information on the dimensions, items and layout of the final questionnaire[10]. Beyond issues regarding content and layout, the values attached to measures such as EQ-5D-5L and EQ-HWB-S are used in decision-making and it is important to ensure that all stakeholders, including members of the public, have an opportunity to reflect and give their views on these values. As the measure is new, it is also important to address its acceptability for use in priority setting. The aim of this study was to gain views from informed members of the public regarding the EQ-HWB-S, including the utility values, for decision-making.

## **METHODS**

The project aimed to ask members of the public about their views of the EQ-HWB-S classifier and the associated utilities. Qualitative methods were considered the most relevant as they would provide opportunity for participants to provide their views. A sample of informed members of the public who were aware of the decision-making process underpinning health technology assessment (HTA) was considered important as this is the context in which the EQ-HWB-S would be used.

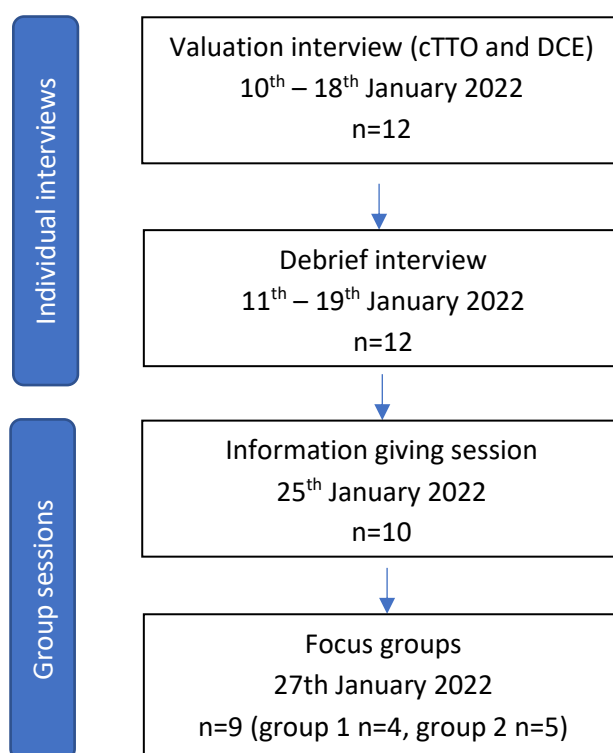
### ***Sampling and recruitment***

Participants were members of the NICE Public Involvement Programme Expert Panel. The Expert Panel is a group of members of the public who are invited to participate in NICE work, including being lay members of HTA Appraisal Committees making them an appropriately informed group. An invitation with an information sheet was sent by NICE in December 2021 where interest could be expressed via an online screener survey. The target sample size was 12 with broad representation of characteristics including age, gender, health, and caring responsibilities. The sample size was considered sufficient to generate different views while also being feasible for online administration. The inclusion criteria were: (1) 18 years of age or older; (2) current UK resident; (3) access to a computer, laptop or large screen tablet with an internet connection and access Zoom or Google Meet; (4) the ability to complete the tasks in English; and (5) able to attend four online sessions on different dates

### **Data collection**

The study comprised four online sessions undertaken in January 2022: (1) a valuation interview; (2) a debriefing interview; (3) a group information/clarification session; and (4) focus groups. The study process is outlined in Figure 1. The first three sessions were aimed at ensuring participants understood the EQ-HWB-S and were familiar with the valuation methodology and how to interpret the utility values. This was considered important to ensure meaningful discussions.

**Figure 1: Flowchart of interviews and focus groups**



The valuation interview replicated the interviews conducted within the feasibility study thereby familiarising the participants with the valuation techniques used to score the EQ-HWB-S. Interviews were completed via computer assisted personal interviews using EQ Portable Valuation Technology (EQ-PVT). All interviews began with four practice cTTO questions, followed by valuing seven states using cTTO and choosing between profiles in seven DCE questions. The states valued were taken from a single block from the feasibility study. The interviews were done by trained and experienced interviewers (n=2) who conducted the interviews for the feasibility study. On consecutive days following the valuation interview, a different experienced qualitative researcher undertook a 1:1 debrief interview (quotes from these are coded using \_DB) with the participants. The debrief was aimed at consolidating participant understanding. They were

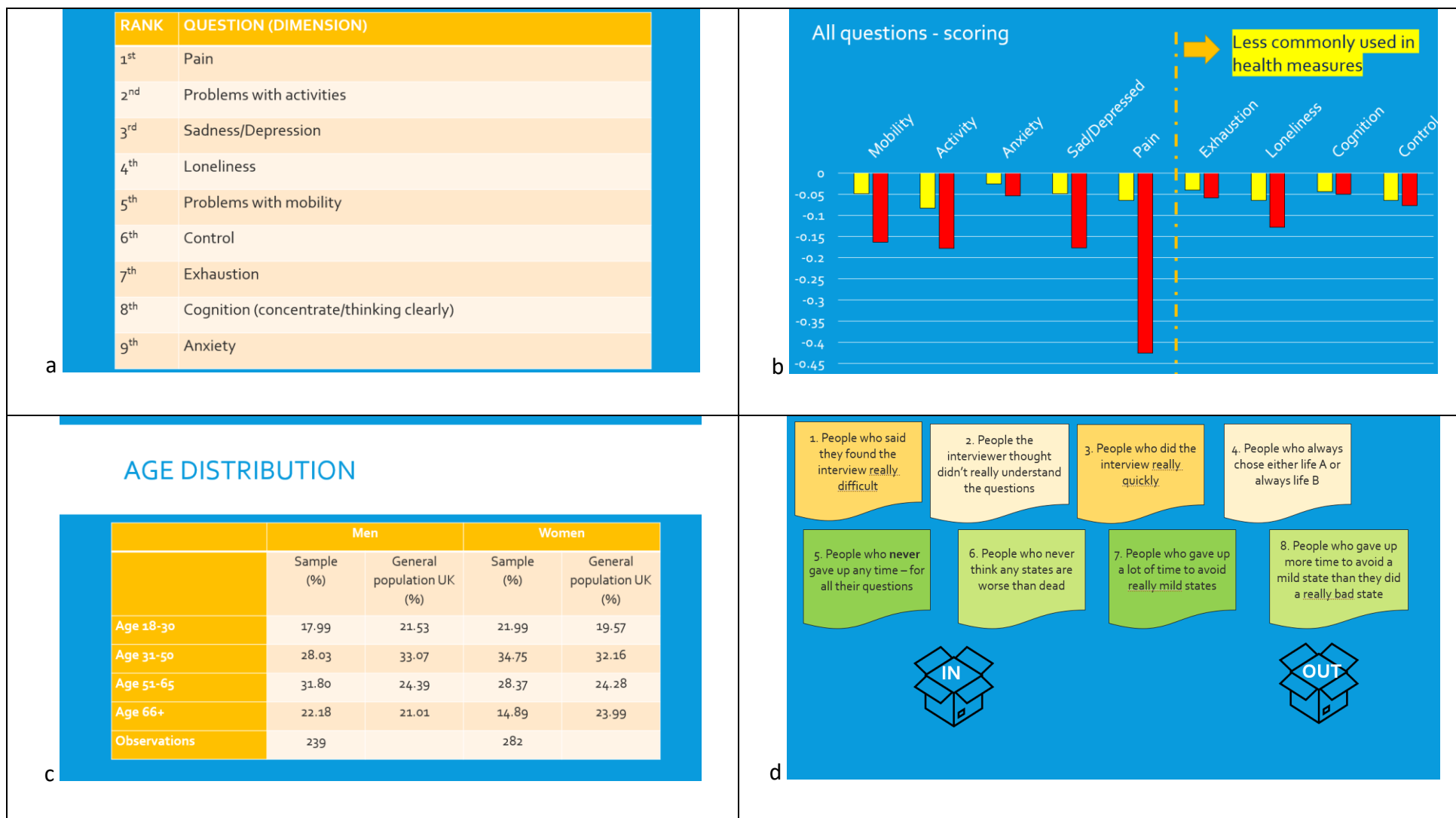
asked about how they found the valuation interview, including questions about the EQ-HWB, cTTO and DCE.

A group information-giving session followed which provided a summary from the debrief interviews, explained the context of the use of EQ-HWB-S in decision-making and reported the values associated with the EQ-HWB-S from the feasibility study with members of the general population. Participants could add further comments and ask for clarification in this session.

The final session was the focus group to elicit views regarding EQ-HWB-S. To facilitate participation for all, participants were split into two focus groups (quotes coded \_FG1 and \_FG2) based on age and gender. Each focus group had an experienced health economist as a facilitator who used a topic guide that was partly informed by previous discussions, an experienced qualitative researcher and a researcher involved in the feasibility valuation study. Participants were presented with results from the feasibility study including the ranking of the dimensions based on the values for the worst response level (Figure 2a), the values (Figure 2b), and information on the sample used including possible exclusions (Figure 2c and d) and asked their views of each with respect to appropriateness for decision-making. To simplify the process, values were only shown for the mid (level 3) and worst (level 5) levels for each dimension (Figure 2b). Participants were also asked about their views on valuation methods from the previous session and how they could be improved.

All sessions took place online via Google Meet or Zoom. Participants completed consent and a background survey (including demographics, EQ-HWB-S, EQ-5D-5L and NICE PPIE experience) online prior to the first interview. Sessions 2 to 4 were audio recorded with full transcription of the focus groups and partial transcription of Session 3 (to capture comments regarding the methods and measure). Participants were offered a financial incentive in line with NICE policy for PPIE reimbursement.

Figure 2 Slides used in the focus groups



### ***Analysis***

Descriptive analysis of the participants' responses to the valuation interviews was undertaken. The audio recordings from the debrief interviews were used to summarise the findings in line with the topic guide questions in order to report these findings in the group information giving session.

Framework analysis[13] was used to assess the findings from the focus groups. Two researchers independently familiarized themselves with the transcripts then generated a coding framework using the topic guide as a starting point which was then applied to the transcripts. This was followed by discussions to confirm, consolidate and reframe the coding framework where necessary. Coding was then undertaken independently again with the new coding framework. Coding was undertaken in MS Word and MS Excel.

Ethical approval for the study was obtained from the University of Sheffield ScHARR Research Ethics Committee (044372).

### **RESULTS**

Twelve participants (out of 48 viable expressions of interest) were selected (50% female across the age range (28 to 74), 75% were white, most were employed (67%) and had a long-term condition (67%) and half had caring responsibilities. All participants had a degree and most reported good to excellent health (72%). Participants were members of the NICE expert panel for an average of 2.7 years (range 1 to 8 years). Participants had prior experience in a variety of roles and activities for NICE including Appraisal Committee lay member and providing feedback on public involvement. All participants completed the first two interviews, two participants did not attend the next two sessions, and one attended all sessions except for the focus group.

#### ***Valuation interviews and debriefs***

In the debrief interviews, participants were generally positive about the EQ-HWB-S. Some dimensions were identified as missing, for example: coping mechanisms, quality of sleep, the burden of caring and relationships. Some aspects were present, but participants felt they were not well-covered. For example, loneliness was considered not to fully capture relationships, aspects of support or being understood, and the physical aspect of cognitive ability was also considered not well covered. There was also discussion regarding time



frame and the variable experience of symptoms (e.g. mental health may vary from day to day or even within the same day). There were mixed views regarding the response options and whether participants could distinguish between them all sufficiently. Participants were aware of EQ-5D from previous NICE roles and noted that adding other quality of life aspects would be useful in the context of decision-making.

The TTO values in this sample spanned the potential range from -1 to 1. Some participants found that the wheelchair practice question was useful, but others were strongly opposed to the choice of example as they considered it to be inappropriate and ableist. This was due to the *“assumption that people in wheelchairs feel that’s a bad thing which they don’t necessarily”* (104\_DB, Female 1, participant 04, debrief interview) and one participant who was a wheelchair user noted that their restriction was due to their condition, not their wheelchair.

Participants considered it difficult to consider the full description of EQ-HWB-S states with aspects that they had not experienced or that they considered to represent unrealistic combinations of problems. Potential ethical issues were also raised by participants, due to a possible emotional impact. One participant noted that the TTO tasks were *“very harrowing and very personally and emotionally affective... I found it a very distressing emotional cycle to go on through the course of the exercise”* (201\_DB, Male 2, participant 01, debrief interview). The debrief interviews were considered to be particularly useful for dealing with the impact of doing the valuation tasks. TTO tasks were also difficult to understand or complex and were considered to be *“trying to be too precise ... I just felt that life is not quite as black and white as that and expecting people to quantify precisely what the trade-off is unrealistic”* (206\_DB). On the other hand, DCE tasks, which did not involve a trade off against years of life, were generally preferred and considered easier to do and understand. However, in some cases, making a choice in the DCE was difficult *“I found that in some of the situations I wanted a wanted a button where they were are equivalent because I didn’t feel that one, living with one set ... were different from another one”* (103\_DB). There were additional comments related to the impact of the characteristics of those taking part (e.g. poor health, experience of different health conditions, age and caring responsibilities) and further discussion was undertaken on this in the focus groups.

## **Focus groups**

Five thematic areas emerged from the analysis of the focus groups.

### 1 Views about the measure

Participants had views on the measure related to the relevance of existing items and whether the EQ-HWB-S covered all that they considered important.

#### 1.1 Views on existing items

##### 1.1.1 Positive views

The inclusion of aspects such as loneliness, control and exhaustion was viewed positively because they would impact on prioritisation of certain types of interventions over and above those related to common physical health dimensions.

*“it would therefore put a stronger focus on the importance of social care because ... they are demonstrating that you can actually make an impact make a difference on all of them but particularly on loneliness and that in combination with some of the other things strengthens the argument for social care community approach” (105\_FG2, female participant focus group 2)*

*“...in my part of the world, the biggest cause of death in men between 18 and 55 is suicide and that investment in pain isn't necessarily going to lead to making any in-roads into suicide figures...” (203\_FG2, male participant focus group 2)*

*“...I'd love to see, some of them anyway, reflected in quality of life impact questionnaires – certainly fatigue as you have exhaustion because that has a massive impact on lives of people who have conditions such as I do and then my other thing what you've got is cognition...” (104\_FG1, female participant focus group 1).*

##### 1.1.2 Combining dimensions

There were suggestions to combine some aspects to make room for others, with the combination of anxiety and depression suggested as something that doctors would consider together: *“I know GPs would look at 3 – depression and 9 [anxiety] together in if you raised these issues with the GP they would do a questionnaire on both so they're very linked and as condition” (205\_FG2).* However, combining was also considered to be potentially

problematic “...anxiety – whether you’re doing that a subservice by adding into sadness / depression” 105\_FG2). The suggestions to combine some dimensions was informed by the knowledge that there was limited room for additional questions in the EQ-HWB-S.

## 1.2 Missing but covered

Participants considered that some questions that had been raised as missing in the interviews and information giving sessions were potentially captured by existing questions e.g. “I think exhaustion to me sort of covers sleep...” (103\_FG1) or “...I thought that could be covered maybe in the everyday activities so dealing with your relationships is part of your everyday activities” (106\_FG2). As with the suggestions to combine, considerations regarding which dimensions were missing was informed by the aim of keeping EQ-HWB-S short enough to be amenable to valuation.

## 1.3 Missing dimensions

### 1.3.1 Coping

Coping was identified as a potentially missing dimension in interviews, the information giving session and in , but this was only discussed in relation to thinking about the state in the trade-offs in one focus group therefore was not considered further.

### 1.3.2 Personal relationships

Although loneliness was linked to relationships, this was not considered to be adequate. Participants did not feel that loneliness reflected the value of personal relationships in terms of support and understanding of carers: “...it’s that carers’ understanding of the condition as well. Those aspects are more important than loneliness. If you haven’t got good support then you’re isolated basically so loneliness is only one little bit of the whole personal relationships...” (204\_FG1). This extended to medical professionals. One participant noted that for some mental health conditions, loneliness was a good thing and they preferred inclusion of personal relationships instead. On the other hand, one participant noted that “...if you just substituted relationships then it wouldn’t define what the problem is and you’d need to be more specific.” (206\_FG1).

## 1.4 Severity rather than frequency

In one group, participants identified that severity may be preferred to frequency as a response option “...all the questionnaires I have been ever asked to do have been how did

*you feel in the last 7 days and I just find that really interesting because perhaps we should be looking at severity not time” (104\_FG1).*

## 2 Explanation for values

Participants views regarding the ranking of EQ-HWB-S dimensions (Figure 2a) and the values (Figure 2b) resulted in four sub-themes as explanations for the values: the interpretation of the language used in the EQ-HWB-S questions and response options, societal norms towards health, respondents’ experience and the impact of the sample.

### 2.1 Interpretation of EQ-HWB-S terms

#### 2.1.1 Interpretation of items

The interpretation of the language used within the EQ-HWB varied and this was considered to have an impact on the values: *“I think part of the problem is language and we all have different definitions in our head ... for me activity is about sport and exercise” (204\_FG1).* For example, the term exhaustion was considered not to capture what it means to experience continuous tiredness:

*“I’d still choose fatigue over exhaustion as its difficult enough for people to understand in the first place but it does at least have the traction in its what commonly used but I think there’s a long way to go to understand fatigue and understand what it actually means” (104\_FG1)*

While at the same time it was considered to be an extreme version of fatigue:

*“It’s the same with exhaustion and fatigue – I would prefer to see fatigue as I understand that. Exhaustion to me is extreme fatigue.” (204\_FG1)*

In addition, some interpreted it to mean burden: *“...for me exhaustion in terms of diabetes feels quite good because it effects your sleep, you get woken up all the time. It’s exhausting because you have to make decisions all day about whether you can go for a walk, whether you need something to eat, how much insulin you have every time you eat but it’s not the same as – what’s [participant’s name]’s thinking is very different to what I’m thinking” (103\_FG1)*

#### 2.1.2 Interpretation of severity levels

There was also interpretation of the levels, for example, the 'moderate' pain level was seen as something which people could cope with: *"...there's a big difference in something you feel you can manage so I'd say most people would say they could manage moderate pain whereas severe pain... putting in the context and asking people to make decisions for trading off life against death, then it would only be in a severe situation that you would start to find that trade off"* (104\_FG1).

## 2.2 Societal norms

### 2.2.1 Norms based on general experience

Values were considered to be a reflection of what was generally experienced as the norm and what was accepted in society. Aspects such as anxiety, lack of control, exhaustion were considered to be the general norm of what people experienced and this had an impact on values *"I think that it did seem to me that certain things that are in the list are probably things that people feel that they already have an experience of and are already parts of their day to day lives so are probably going to be lower down on that ranking"* (203\_FG2). This participant also raised a link between high levels of deprivation and perceiving problems these areas as the accepted norm.

### 2.2.2 Current societal views on what is important

Participants in one group considered that the values reflected what was important in society *"...I think society values things that are physical much more than things that are mentally exhausting or lack of control"* (103\_FG1).

## 2.3 Respondents Experience

### 2.3.1 Knowledge or experience

The lack of direct experience of the states being valued of those taking part was considered to impact on how they valued certain aspects. *"...in your sample how many people would you have had that will have suffered from any of the conditions that I've mentioned such as stroke or long covid or conditions that lead to extreme fatigue and therefore have an understanding of what that is other than feeling a bit tired which is what most people of it as."* (104\_FG1)

Conversely, common experience of mild levels of certain aspects was considered to dilute the understanding of what it meant to experience those aspects in a way that was harmful. *"I definitely don't think anxiety should be as low as it is and I completely agree with what [other participant] said there's a natural state of being for a lot but those who do suffer from it or have it as a real core part of who they are I'd strongly posit that they'd put that much higher"* (201\_FG2). These commonly experienced aspects such as anxiety or being tired in general were considered to have an impact on how individuals valued those aspects in the EQ-HWB-S.

### 2.3.2 Characteristics of participants and 2.3.3. Experience of external factors

Beyond experience of the aspects covered in the EQ-HWB-S, other characteristics of respondents were considered important such as age *"...We've talked about patient population – certainly as you get older those things matter more"* (205\_FG2) and their financial status *"...I think if you're looking at factors that are going to have a big play on people's take on these things then whether they're working, in work, poverty, financial problems and general poverty and the grind of poverty..."* (203\_FG2). These were considered as having a direct impact on knowledge or experience which then impacted on how states were valued.

## 2.4 Impact of the sample

The sample used in the feasibility valuation survey was considered to have a direct impact on the ranking and values that were presented to the focus group members.

### 2.4.1 Size

Separately from the knowledge and experience of those taking part, values were considered to be impacted by the size of the feasibility study sample (n=521) which may also interact with experience/knowledge: *"if you don't have the experience then that comes back again to the sample size and that you're not going to pick these people up"* (104\_FG1).

### 2.4.2 Composition

The composition of the sample that undertook the survey was considered to have a direct impact on the values and ranking. For example, *"... you're not going to get a picture of ethnic communities by having a sample that is 82% white"* 104\_FG1). The proportion of the

feasibility study that was older was considered to be likely to have an impact, especially the lower proportion of women aged 66 and above.

### 2.4.3 Pandemic

The feasibility valuation survey was undertaken during the pandemic and focus group members felt that this would have had an impact on the sample and consequently the values. *“Exhaustion again I’m not surprised that’s quite low down - that’s a state a lot of people especially in the last few years will have experienced because of the impact covid would have had on the wider populous and therefore that state of being continually being exhausted and that feeling of no control is very much the state that a lot of people have viewed so if that’s the context in which this survey was carried out”* (201\_FG2).

## 3 Concerns about values

### 3.1 Appropriateness of values

Participants had different concerns about the values that fell under three sub-themes: appropriateness of the values, subjectivity and the lack of difference between levels.

#### 3.1.1 No surprise related to ranking

Participants were not surprised by the high ranking of pain and they did not express any concerns about this dimension having the largest weight. *“I don’t think that anybody would argue with pain as being really the most serious”* (206\_FG1)

#### 3.1.2 Surprise related to ranking

On the other hand, there was surprise about the ranking of anxiety: *“I definitely don’t think anxiety should be as low as it is”* (201\_FG2) as participants felt that this was an important aspect of mental health and especially given the increased anxiety due to the pandemic: *“I’m surprised anxiety is at the bottom as the pandemic has caused huge amounts of anxiety so the fact it’s down there does surprise me.”* (204\_FG1). There was also surprise about the ranking of exhaustion: *“... exhaustion does cover the burden of that but actually its very intrusive in terms of being able to do normal things but it is ranked very low which is also surprising”* (103\_FG1) but at the same time, one participant noted that *“Exhaustion again I’m not surprised that’s quite low down - that’s a state a lot of people especially in the last few years will have experienced because of the impact covid...”* (201\_FG2) Control was also surprising *“I think I was surprised that control was further down because I felt it was a quite*

*a broad measure but if you felt that you had had some control over your life then you might be feeling a bit more well so when I was doing the exercise I felt that I did look to that one"* (103\_FG1).

### 3.2 Arbitrariness and subjectivity of ranking

One participant noted that apart from pain, the rest of the dimensions could be in any order: *"The rest of it feels fairly arbitrary. ... But the rest of it I could see in almost any order."* (104\_FG1) and another that there was an element of subjectivity related to ranking these dimensions *"...I think what we're seeing is actually the difficulty in ranking all of these and how very subjective it is. ... the ranking of the others [apart from pain] – I don't buy into the fact that you can really rank them even if you take a very large sample of people and try to score it."* (206\_FG1).

### 3.3 Lack of difference between levels

Apart from the ranking, there were concerns regarding the lack of differences between severity levels (levels 3 and 5) for cognition and control *"I was most surprised by was moderate and severe bars in cognition and control – I would have thought that the less or more severe lack of control or cognition, individuals experience the impact as being far more significant and to see only moderate changes between level 3 and level 5 response"* (201\_FG2).

## 4 Views on who to include

### 4.1 Who to include?

Participants had strong views about who should take part in valuation studies. Factors that were considered to be important in selection included *"...various disabilities"* (106\_FG1) with specific mention of sensory disabilities and mental health, *"...age especially in a country where the population is living longer"* (201\_FG2) and *"..... socioeconomic and regional factors..."* (203\_FG2). Socioeconomic factors included ethnicity, sexuality, faith and deprivation. Taking into account *"...urban and rural deprivation..."* (105\_FG2) was considered important as rural deprivation was linked to different experiences to urban deprivation. There were concerns raised about potential participants (sometimes referred to as 'hard-to-reach' but this may reflect recruitment strategies) who would not engage with this type of research. There were recommendations to use disability panels or to take time



building relationships to allow engagement with any groups identified as such, and not just relying on community leaders to access communities.

## 4.2 How big should the sample be?

### 4.2.1 Absolute size

The absolute size of the sample was also highlighted as an issue as it allowed different groups identified for inclusion to be taken into account *"...you'd have to look at this across those different groups with a much larger sample size"* (104\_FG1). It was also considered to be important based on the population size *"...So yes while its agreeable that this group [feasibility study sample] has a really good distribution based on the general UK population of a wider UK context if we were looking at maybe a 1% or 2% [of 67 million] ..."* (201\_FG2) while acknowledging that increasing the sample size to that level would be challenging.

### 4.2.2 Relatively larger for under-represented or key groups

There were also suggestions to over-represent certain groups to allow adequate representation *"My suggestion is that you'd use these groups as boosted samples so that you wouldn't be doing it as a nationally representative sample ...but if you have a nationally representative sample you are never going to represent the groups properly"* (104\_FG1). This was highlighted as a way to ensure that the values were reflective of different groups rather than just large groups.

There was some discussion regarding having a nationally representative sample and participants understood that *"you can't just increase the number because then you wouldn't have a nationally representative sample but you can take a separate boosted sample and then compare that to your nationally representative sample"* (104\_FG1).

## 4.3 Exclusions

The focus group participants gave either conditional support or no support with regards to criteria for possible exclusions that were presented to them.

### 4.3.1 Conditional support for exclusions

Exclusions from respondent data were considered problematic unless a lack of understanding or engagement could be confirmed that was not related to an interviewer doing their task well e.g. not highlighting inconsistencies. Relying on more than one source

of information to make a judgement about whether participants understood or engaged was considered more helpful e.g. combining how quickly they completed the tasks, with selecting just one side in the DCE profile and the interviewer observing that they did not engage or understand.

#### 4.3.2 No support

There was agreement that respondents who found the interview difficult, completed it quickly, who did not identify any states as worse than dead or who did not give up any time at all should not have their data excluded. There were two different reasons why focus group participants did not want to support exclusion of data.

##### 4.3.2.1 Reflect true preferences

One was that responses could reflect true preferences *“It would be really difficult position for us to be in to disregard any of these things [list of reasons for exclusion] because everyone’s entitled to their own perceptions and ways of interpreting and dealing with the survey.”* (201\_FG2) There was also concern that what could be interpreted as an inconsistency i.e. giving up more time for a mild state compared to a severe state, may reflect a rational choice *“I don’t know, they might have things in their mind about why they’ve made that decision that if they’re in a really bad state they feel like someone’s caring for them whereas if it was a mild state, they have to deal with that themselves. There might be some reason in their mind why they’ve made that decision and it is their opinion isn’t it. I think as long as the interviewer ... doesn’t think that they’ve misunderstood the task, then I think that is their view isn’t it?”* (103\_FG1). One respondent summarised the need to reflect individual preferences saying *“you’re asking for people’s opinions and as long as it’s pointed out to the that there may be a contradiction in what they say then you have to accept their answer. Everybody has a different priority. You cannot impose the interviewers’ priorities onto the interviewee.”* (206\_FG1)

##### 4.3.2.2 Moral or ethical issues

There were concerns regarding the ethics of excluding data especially after participants had gone to the trouble of engaging with challenging tasks. *“... I’m a bit weary of the principle of researchers jettisoning information. I’m sure it would be done honourably but you just never*

*know. If you disregard this, what else might you disregard so there's a bit or a moral issue there."* (105\_FG2)

## 5 Difficulties with TTO/DCE tasks

There were difficulties with completing the TTO and/or DCE tasks that were related to commonly encountered problems with valuation tasks such as difficulties with separating related dimensions:

*"... I find it difficult at times to separate things out so that if loneliness is a big problem, that might also impact on your depression and anxiety and your control so it's hard to think of them as separate things...."* (106\_FG2);

issues with trading-off life years

*"...I just think it's very difficult to someone to think I will give up some of my life because unless the circumstances are very extreme..."* (104\_FG1);

and an inappropriate length of life

*" ... TTO was quite hard because it's 10 years which doesn't seem very long whereas if it was over a sort of lifetime you might think I would be willing to give up a couple of years if I had this lifetime..."* (103\_FG1).

DCE was preferred to TTO but participants made recommendations on how to improve the overall valuation exercise via providing support.

### 5.1 Increase introduction

Participants recognised the value of TTO for scoring but agreed that expanding the introduction would be helpful. This was included providing reasons for doing the exercise *"...a bit more lead in, more information, more context and also thinking I know a little bit about QALYs ... from my experience with NICE about what you're trying to do and I think for someone without any of that background, to go into it in the way we did is like pushing someone who can't swim into the end of a swimming pool..."* (203\_FG2).

It could also be used to prepare them for the potentially emotional impact to mitigate any risk of unexpected emotional impact *"....When you speak to people who live with disabilities*

*or long term health conditions you are going to use current word 'trigger' people quite substantially..." (104\_FG1).*

There was acknowledgement that providing more information up front had the potential to discourage participants in particular groups *"...but at least you've asked them, you've given them the option and that's hugely important and also the interviewers need to capture... if you ask a group ... and no one wants to take part then that's something you need to be aware of that's something to note in itself"* (204\_FG1).

## 5.2 On-going support

On-going support from the interviewer was considered important for TTO *"...because of the complexity and the nuisance of trying to get the trade-off accurately done..."* (201\_FG2). This included getting interviewers to highlight inconsistencies when they occurred in the interview.

## DISCUSSION

This study gained the views of informed members of the public regarding the EQ-HWB-S, a new measure of health and wellbeing that can be used to support decision-making in health and social care including for informal carers. Generally, participants had positive views about the additional wellbeing dimensions in the EQ-HWB-S and although they identified some missing dimensions, apart from personal relationships there were none that they considered to be completely missing. Participants were surprised by the ranking and the values for some of the additional dimensions, but they highlighted a number of reasons why these might be the case. They raised concerns about the sample and the valuation methods used.

### ***What did we learn that we did not know?***

Participant views regarding the valuation tasks were useful for us to reflect on. As researchers used to undertaking valuation surveys, we had not reflected on issues such as the 'wheelchair example' or the way in which the TTO tasks can have an emotional impact on some participants. All participants received an information sheet that highlighted the use of comparisons and that these would include reference to death in cTTO but this was not sufficient to prepare participants for the 'die immediately' option in the TTO. We recommend further considerations of what and how information is provided prior to

interviews, consideration of an appropriate initial practice state and how to provide support after the interviews for those who want it. **Potential discussion questions:** Is further research to assess whether the ‘wheelchair’ example is acceptable as a practice question given the issues identified relating to ableism and its focus on the intervention (wheelchair) rather than the state (mobility)? What practice tasks could be used instead, and what research would be needed to support their use? What responsibility do we have as researchers in preparation for and providing support following potentially difficult valuation interviews?

The participants had strong views regarding representation across different groups when recruiting for valuation studies. This included representation on aspects such as urban and rural deprivation to ensure that values would represent those who would be impacted by use of the EQ-HWB-S. The use of public engagement groups to inform sampling prior to data collection can inform on these type of issues.

The participants also had strong views about exclusions after data collection. There was limited support for exclusions when this relied on a single criteria including some that we may consider to be more objective such as inconsistencies.

### ***What did we learn about engaging with informed members of the public?***

One of our main aims was to assess the acceptability of the values for use in priority-setting. This was important as we know that the dimensions of the EQ-HWB-S matter to patients, informal carers and social care users care but it is not used with equally weighted scoring across these dimensions. It is therefore valuable to subject utility values to scrutiny from informed members of the public, as agreement on the questionnaire alone does not ensure acceptability. However, there is a question as to whether a small group can be used to represent public views which participants were aware of “...we thought that anxiety should be higher up the scale but that’s just our opinion we’re only two individuals not the whole sample and some people may say anxiety is down where it should be.” (206\_FG1). **Potential discussion question:** Should utility values be subject to acceptability assessment? If so, who should provide views of acceptability?

In order to ensure that everyone had the same starting point, there were a number of sessions which required resources. Despite the number of sessions, it would have been valuable to have had more time to cover specific issues regarding the impact of using values with concrete examples from a decision-making context and methods of sampling survey respondents. Preparing for this type of work with members of the informed public as researchers may have highlighted some of these issues prior to running the sessions.

### ***Strengths and limitations***

As members of NICE Expert Panel, participants were all highly motivated to contribute and had encountered other measures such as EQ-5D so understood the context in which EQ-HWB-S would be used. However, some of this knowledge may have impacted on what they discussed e.g. suggestions to combine anxiety and depression based on their knowledge of EQ-5D. Furthermore, we did not have any PPIE partners to inform this stage of the project; the larger project PPIE group used in previous stages were service users who did not have HTA experience.

Participants undertook the valuation interview in the exact format, with the same preparation and the same interviewers as was done in the main feasibility study; they were also supported to subsequently reflect on the exercise and gain a bigger picture of the purpose of the measure. This gave them a good level of understanding of the EQ-HWB-S and the utility values and a unique vantage point to recommend practical and ethical improvements in how elicitation exercises are conducted, and what data should be included in the analysis. However, undertaking the first three sessions led to some necessary repetition and had a time implication.

The online engagement enabled participants to engage without geographical restrictions, but may have limited access for those who would have preferred face-to-face engagement. The sample size was small and only two focus groups were undertaken. We aimed to represent different characteristics but loss of some of the participants during the study meant this was not fully achieved.

### ***Conclusion***

The EQ-HWB-S was viewed positively by informed members of the public but there were concerns regarding the valuation methods, sample and resultant values. We have made

recommendations in relation to valuation and the inclusion of informed members of the public in this type of research.

## REFERENCES

1. Brazier, J., et al., *The EQ Health and wellbeing: overview of the development of a measure of health and wellbeing and key results*. Value in Health, 2022.
2. National Institute for Health and Care Excellence, *NICE health technology evaluations: the manual*, NICE, Editor. 2022.
3. Finch, A.P., J.E. Brazier, and C. Mukuria, *What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews*. The European Journal of Health Economics, 2018. **19**(4): p. 557-570.
4. Netten, A., et al., *Outcomes of social care for adults: developing a preference-weighted measure*. 2012. **16**: p. 16.
5. Peasgood, T., et al., *Developing a New Generic Health and Wellbeing Measure: Psychometric Survey Results for the EQ Health and Wellbeing*. Value in Health, 2022.
6. Peasgood, T., et al., *Criteria for item selection for a preference-based measure for use in economic evaluation*. 2021, Springer. p. 1425-1432.
7. Mukuria, C., et al., *Qualitative Review on Dimensions of Quality of Life Important for Patients, Social Care Users, and Informal Carers to Inform the Development of the EQ-HWB*. Value in Health, 2022. **25**(4): p. 492-511.
8. Carlton, J., et al., *Generation, Selection, and Face Validation of Items for a New Generic Measure of Quality of Life: The EQ Health and Wellbeing*. Value in Health, 2022.
9. Mukuria, C., T. Peasgood, and J. Brazier, *Applying EuroQol Portable Valuation Technology to the EQ Health and Wellbeing Short (EQHWB-S): a pilot study*. School of Health and Related Research, University of Sheffield Discussion Paper Series, 2021.
10. Carlton, J., et al., *The role of patient and public involvement and engagement (PPIE) within the development of the EQ Health and Wellbeing (EQ-HWB)*. Journal of Patient-Reported Outcomes, 2022. **6**(1): p. 1-8.
11. Wiering, B., D. de Boer, and D. Delnoij, *Patient involvement in the development of patient-reported outcome measures: The developers' perspective*. BMC health services research, 2017. **17**(1): p. 1-10.
12. Carlton, J., et al., *An emerging framework for fully incorporating public involvement (PI) into patient-reported outcome measures (PROMs)*. Journal of Patient-Reported Outcomes, 2020. **4**(1): p. 1-10.
13. Ritchie, J. and L. Spencer, *Qualitative data analysis for applied policy research*. The qualitative researcher's companion, 2002. **573**(2002): p. 305-29.