University of Kent

Kent Academic Repository

Brazier, John, Peasgood, Tessa, Mukuria, Clara, Marten, Ole, Kreimeier, Simone, Luo, Nan, Mulhern, Brendan, Pickard, A. Simon, Augustovski, Federico, Greiner, Wolfgang and others (2022) ***The EQ Health and Wellbeing: Overview of the Development of a Measure of Health and Wellbeing and Key Results.*** **Value in Health . ISSN 1098-3015.**

Downloaded from

https://kar.kent.ac.uk/93584/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1016/j.jval.2022.01.009

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Title: The EQ-HWB: overview of the development of a measure of health and well-being and key results

Authors: John Brazier (PhD)[1], Tessa Peasgood (PhD)[1, 10], Clara Mukuria (PhD)[1], Ole Marten[2], Simone Kreimeier (PhD)[2], Nan Luo[3], Brendan Mulhern (PhD)[4], A Simon Pickard (PhD)[5], Federico Augustovski (PhD)[6], Wolfgang Greiner (PhD)[2], Lidia Engel (PhD)[7], Maria Belizan (MSc)[6], Zhihao Yang[8], Andrea Monteiro[5], Maja Kuharic[5] , Luz Gibbons[6], Kristina Ludwig (PhD)[2], Jill Carlton (PhD)[1], Janice Connell[1], Stacey Rand (PhD) [9] , Nancy Devlin (PhD)[1,10], Karen Jones (PhD)[9], Aki Tsuchiya (PhD) [1], Rosemary Lovett (PhD)[1,11], Bhash Naidoo[11], Donna Rowen (PhD) [1], Juan Carlos Rejon-Parrilla[11,12]

[1]School of Health and Related Research, University of Sheffield, UK

[2]Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Germany

[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

[4]Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia

[5]Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Illinois, USA

[6]Institute for Clinical Effectiveness and Health Policy (IECS-CONICET), Buenos Aires, Argentina

[7]Deakin Health Economics, School of Health and Social Development, Deakin University, Australia.

[8]Health Services Management Department, Guizhou Medical University, China

[9]Personal Social Services Research Unit, University of Kent, UK

[10]Melbourne School of Population and Global Health, University of Melbourne, Australia

[11]National Institute for Health and Care Excellence (NICE), UK

[12] Andalusian Health Technology Assessment Unit (AETSA), Spain

Corresponding Author: John Brazier, j.e.brazier@sheffield.ac.uk. 44-(0)114-2220726

School of Health and Related Research, Regent Court, 30 Regent Street, University of Sheffield, S1 4DA,UK

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

Disclosures: 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, the Department of Health and Social Care or those acknowledged.

Acknowledgments: We would like to thank NICE for highlighting the methodological research need to the MRC that resulted in the funding call entitled "Beyond the QALY" which led to this research being funded. We would like to acknowledge the support of the National Institute for Health Research Yorkshire and Humber Applied Research Collaboration (formerly CLAHRC) and the National Institute for Health Research Clinical Research Network (NIHR CRN). We acknowledge the invaluable contributions of members of the project steering group, advisory group and public and patient involvement and engagement groups and Julie Johnson for project administration. We would also like to thank members of the EuroQol Group Association for their input at plenary and academy meetings and the EuroQol office for their support. Finally, we would like to acknowledge the contribution of all the patients, social care users and informal carers who took part in all the studies across the different countries.

Conflicts of interest: John Brazier, Tessa Peasgood, Clara Mukuria, Ole Marten, Simone Kreimeier, Lidia Engel, Zhihao Yang, Kristina Ludwig, Brendan Mulhern, A. Simon Pickard, Nan Luo, Federico Augustovski, and Wolfgang Greiner are members of the EuroQol Group Association

Keywords: Outcome measures, QALYs, Utilities, health and wellbeing, EQ-HWB, preference-based measures, social care outcomes, carer outcomes

Highlights:

1. Measures for estimating Quality Adjusted Life Years (QALYs) attempt to capture the health of patients but miss broader wellbeing considerations such as autonomy and relationships that are important to patients and care users. Furthermore they do not consider the impact on carers' quality of life.
2. This paper presents an overview of the international development of a new generic measure, the EQ Health and Wellbeing (*EQ-HWB*) and a short version the EQ-HWB-s using qualitative evidence from service users (health and care services) and carers, along with psychometric evidence, collected in six countries.
3. EQ-HWB is designed for evaluating interventions in health care, public health and social care with the intention that it can be used to estimate QALYs to inform decision making both within and between these sectors.

Abstract:

Objective: Existing measures for estimating Quality Adjusted Life Years are mostly limited to health-related quality of life. This paper presents an overview of the development the EQ-HWB, which is a measure that encompasses health and wellbeing.

Methods: Stages: i) Establishing domains through reviews of the qualitative literature informed by a conceptual framework. ii) Generation and selection of items to cover the domains. iii) Face validation of these items through qualitative interviews with 168 patients, social care users, general population and carers across six countries (Argentina, Australia, China, Germany, UK, US). iv) Extensive psychometric testing of candidate items (using classical, factor analysis and item response theory (IRT) methods) on over 4,000 respondents in the six countries. Stakeholders were consulted throughout.

Results: 32 sub-domains grouped into seven high level domains were identified from the qualitative literature and 97 items generated to cover them. Face validation eliminated 36 items, modified 14 and added three. Psychometric testing of 64 items found little difference in missing data or problems with response distribution, the conceptual model was confirmed, and most items performed well in the IRT in all countries except China. Evidence was presented to stakeholders in two rounds of consultation to inform the final selection of items for the EQ-HWB (25-item) and the EQ-HWB-S (9-items).

Conclusion: EQ-HWB measures have been developed internationally for evaluating interventions in health, public health and social care including the impact on patients, social care users and carers.

1. INTRODUCTION

The Quality Adjusted Life Year (QALY) has become a widely used measure of outcome for use in informing decision making in health technology assessment (HTA).[1] QALYs provide a way to capture benefits in terms of impact on survival and health-related quality of life (HRQoL) valued on a utility scale where one is full health and zero is equivalent to being dead. Widely used measures for estimating HRQoL on this utility scale are the EQ-5D, SF-6D and HUI3.[2-5] These measure an individual's general health; for example, the EQ-5D describes health across five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression, with three or five levels,[3,6] with an accompanying utility value set usually obtained from the general public.

The most widely used health measure in economic evaluation, the EQ-5D has been found to be valid in many medical conditions, but performed poorly in others e.g. dementia, multiple sclerosis, severe and complex mental health, and hearing and mixed results in vision.[7] There is also evidence that medical conditions impact important outcomes beyond health.[8] Furthermore, resource allocation decisions are made in related sectors of personal care and other forms of practical support (known in the UK as social care) and public health, which would benefit from broader utility measures to calculate QALYs. Importantly, the provision of care may not only improve HRQoL, but also broader aspects of quality of life (QoL) for the recipients and potentially their carers from better meeting their wants and needs in terms of social care (e.g., improved relationships, greater independence and control). There are also important consequences for the QoL of family or friends who care for them and increasingly there are interventions designed to support them (e.g., respite care).

Health measures have limited ability in capturing important outcomes in other sectors. As a result, the Adult Social Care Outcome Tool (ASCOT) has been developed for use in social care in England.[9] For the impact on carers, there are measures like CarerQoL-7D, and ASCOT-Carer that have been developed for use in economic evaluation.[10,11] This proliferation of measures does not allow comparison across sectors. Furthermore, many interventions have impacts across sectors (e.g., providing a meal delivery service improves health), but using health and social care outcome measures together risks double counting and using a sector specific measure will miss important outcomes. Developing a common outcome measure, suitable across these sectors, will provide better evidence to strengthen cross-sector decision making.[12]

There are number of instruments available that could provide a common measure to be used across sectors. There are subjective wellbeing (SWB) measures, including single self-reported items on happiness and life satisfaction. There are also multi-items measures such as the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) which contains 14 positively worded items to measure concepts of emotional feelings and psychological flourishing.[13] Currently these measures do not provide metrics that could be used for estimating QALYs. There has been work to develop monetary valuations of improvements in SWB, but these are not well developed and monetary valuations are not typically used in health care.[14] Other generic measures that could be used across sectors include the ICECAP measures that has been developed to assess a person's capability in terms of what they 'can do' and 'able to be' covering aspects related to attachment, stability/security, achievement/role, enjoyment, and autonomy/control.[15,16] This is based on Sen's capabilities framework that is concerned about what a person can do as well as what they actually achieve in terms of functioning.[17] Both ICECAP-O for older people[15] and ICECAP-A for adults[16] have preference-based scores estimated form

general population values elicited using Best Worst Scaling. They have not been anchored on the zero to one scale used to estimate QALYs, although they can used in some forms of economic evaluation.

An important issue is whether decision makers want to rely exclusively on either SWB measures, health measures or some combination. This issue was explored in a qualitative study of the role of wellbeing in resource allocation decisions in health and social care with decision makers in the UK undertaken in preparation for this research.[18] There was broad range of opinions about the role of wellbeing. One broadly held view was that there was a need for improved consideration of broader quality of life outcomes. However, there was considerable caution in relation to the use of SWB alone due to concerns over interpersonal comparability.[18] Some decision makers, particularly in HTA wanted to retain the primacy of health to measure benefits. This work influenced the decision to look to develop a measure that combined health and wellbeing domains to enable comparison across sectors.

For this reason, we embarked on a large, international study to develop a new generic measure of health and wellbeing, the EQ Health and Wellbeing™ (EQ-HWB) that has the potential to estimate QALYs for use across health and social care settings. This paper provides a high-level overview of the development of the EQ-HWB. It starts with some background to the study before describing the stages involved, primary results and how these were used in the development of the EQ-HWB. The reader is provided with references to the accompanying articles in this special issue and elsewhere that provide more in-depth detail.

2. Background

2.1 Theoretical approach

The theoretical approach underpinning EQ-HWB fits within the extra-welfarist tradition both in aiming for a multi-dimensional measure of benefit and in using social preferences to judge the value of states.[19] As Culyer has argued the 'characteristics of people' should be taken into account in addition to utility when evaluating healthcare.[20] The 'characteristics of people' includes whether they are happy, pain free, physically mobile, free to choose and so forth. Furthermore, we wanted a measure that met three requirements.[21] Firstly, the new measure covers aspects of quality of life that have been identified as important by service users and their carers. Secondly it meets predefined criteria based on being fit for purpose and able to inform resource allocation published in detail elsewhere, including being a single index measure anchored on the zero (equivalent to dead) to one scale (equivalent to full quality of life) which reflect social values, have a reasonable degree of confidence in it having interpersonal and inter-temporal comparability, amenable to routine use and clinical trials and have good measurement properties. The third requirement is that it should be in line with what policy makers in health and social care think is important to their decision making.

2.2 Conceptual foundation of EQ-HWB

The development of the EQ-HWB was informed by an initial framework based on the widely known Wilson and Cleary's model of HRQoL.[22] This model links biological and physiological variables, symptom status and then

to functional health, health perception and finally quality of life. It was modified to include aspects beyond health, with biological functioning extended to include other factors that may have a direct or indirect impact on physical and mental health symptoms, such as being a carer (Figure 2). Impact was extended beyond functioning and activity to include broader aspects of quality of life, including control and identity. The Wilson and Clearly framework was chosen because it provided a simple structure for direct and indirect impacts of circumstances and physiological variables and a useful way of thinking about the connection to environmental and individual mediators, without over-imposing model complexity that would be incompatible with our main approach. More detail and rationale for this framework has been published elsewhere.[19]

2.3.Stakeholder involvement

A patient and public involvement and engagement (PPIE) group participated directly in the project throughout the development of the EQ-HWB. This has been detailed elsewhere.[23] The group consisted of seven members with diversity in gender, ethnicity, and health condition representation across the group. Discussions from each PPIE group session and the outcome of any tasks they undertook were shared with the wider research team and used to help inform decision-making.

There was an online Advisory Group (n=124) with members of HTA agencies, broader health and social care commissioners and academics. Furthermore, members of the National Institute of Health and Care Excellence (NICE) have been a part of the project team (RL, BN, JCRP) while other NICE staff were invited to provide feedback at different stages of the project. Workshops with members of NICE and the NICE Citizens Council were also held at different stages of the project. Lastly, there was project Steering Group (n=12) made up of academics and stakeholders to oversee the scientific work, the processes of stakeholder engagement and provide feedback throughout the project.

OVERVIEW OF THE STAGES OF DEVELOPING THE EQ-HWB

The five stages of the development the EQ-HWB measures are summarised in Figure 1.

Stage 1: Establishing the domains

A review was undertaken of the qualitative literature of the way patients, social care service users and carers (over the age of 18) report that their conditions, circumstances and related interventions, impact on their QoL. The methods and results are detailed in a companion paper.[24] The focus was on qualitative research using methods such as interviews and focus groups directly undertaken with them published up to October 2017. The qualitative review included: 1) a targeted sample of qualitative reviews in 15 conditions chosen to represent a broad range of physical and mental health problems; 2) qualitative literature concerned with the development or testing of generic measures designed for the evaluation of interventions in health, social care and/or carers. Framework analysis and synthesis was undertaken based on the conceptual model (Figure 2).

The literature review resulted in 32 sub-themes grouped into seven high level themes (Feelings and emotions. Cognition, self-identity, autonomy, relationships, physical sensations, activity) (Figure 3). Aspects related to

'feelings and emotions', 'activity' and 'relationships and social connections' were recurring themes across the different populations, including those with physical and mental health conditions. These high-level themes were similar although not identical to the initial conceptual framework in Figure 2.

Stage 2: Preliminary selection of items

A comprehensive list of candidate items was generated and categorized into the domains and sub-domains from stage 1 (n=687) as detailed in a companion paper.[25] There were several iterations of item reduction following group sessions with members of NICE Citizens Council, the PPIE group and a survey of the advisory group. These groups were given a set of item selection criteria based on existing published criteria[26,27] and adapted following consultation with the Steering and Advisory Groups (for details of consusltation see[21]). This stage led to some modification of the domain structure e.g., combining sub-domains reflecting control and autonomy, and removal of 'burden' as a sub-domain. Further modification of items was made following the consultation and none of the 97 items covering 28 sub-domains were taken forward to face validation in their original form.

A seven-day recall period was chosen for the items based on competing considerations. Asking about today risks missing important events that do not occur every day, such as episodic conditions like asthma, unless data are collected every day. Furthermore, it is difficult to frame items for sub-domains like control, coping, loneliness and doing the things you want to do around a single 24-hour period. By contrast, a respondent may struggle to remember their experiences over the last month.

Alternative response choices (including frequency, difficulty and severity) worked differently across the items, so a selection of response choice options for each item was used in the face validation to get the views of participants.

Stage 3: Face and content validation

This stage explored the face validity of a the set of potential items and is detailed in a companion paper.[25]

Samples were recruited to adequately cover all key interested groups across the six countries and to ensure that each proposed item was considered by sufficient sample of people. Each participant saw a subset of the items (about 40) from between 2-3 domains to avoid excessive burden. For each item, respondents were asked to think about their response to the item, what they thought about when they read the item, and how easy they found it to understand and answer. In some cases, several different response options (such as frequency versus intensity) were presented. This study also examined the cross-cultural relevance of the item content, interpretation of the wording, and translation issues arising from the forward and backward translation of the items into three languages of Spanish, German and Chinese. This information was combined to inform decisions about which items to drop, take forward with or without refinement and response options for the psychometric survey and the final item selection in stage 5.

There were some common findings across the groups interviewed and the six countries. Participants preferred simpler layouts, but they wanted more information on context or definitions to help them complete

the items. Participants did not have a clear preference between the positive or negative framing of items or alternative response options, or preferences were expressed without any clear reasons for them. Some items worked well for one group but not as well for another group e.g., being able to look after yourself was an issue for patients but not carers. Participants often forgot or ignored instructions e.g., the recall period.

These findings resulted 64 items being taken forward (47 with no amendment, 14 with modification, 3 new) covering 26 of the sub-domains identified in the literature review and one new additional sub-domain (self-confidence) and one dropped (dignity/respect). Explanations for the decisions are shown in supplementary materials to a companion paper.[25]

Stage 4: Psychometric study

The aim of this stage was to examine the domain structure, and test the psychometric performance of candidate items. Details are reported in a companion paper.[27] Patients, social care users, carers and members of the general population were recruited in the UK and in different proportions across the population groups in the six countries. Recruitment was online in all countries with an additional paper-based sample in the UK. This resulted in 4,830 participants being included in the psychometric analyses across the six countries.[28]

Dimensionality was examined by factor analysis. The conceptual model was reasonably well confirmed, particularly across UK, Australia, US, Argentina and Germany, with some modification in the latter two countries, namely, dropping energy and combining mobility and self-care in Germany and combining self-worth and coping in Argentina. There was evidence of high correlation between factors but combining them did not improve overall model fit. The model did not fit the China data as well where many of the feelings sub-domains needed to be combined.

In terms of item performance, there was little difference in rates of missing data across items (ranging 0.6 to 6.4%). The distribution of responses did not suggest floor or ceiling effects that could not be justified by the prevalence of the condition (e.g., poor hearing has a comparatively low prevalence in the general population). Most items were able to distinguish between those with physical and mental health conditions as well as by severity where this was tested. However, evidence on known group differences was mixed for carers since the caring role can make health and wellbeing worse (e.g. from feeling useful) or worse (e.g. feeling tired).[29] Item performance was also assessed using Item Response Theory (IRT) which indicated that the response levels of most of the items were ordered, but items functioned differently across some groups. Overall, 32 items worked well, 25 had mixed evidence and seven performed poorly. This evidence was taken forward to the next stage.

Stage 5: Selection of items for the measures

The aim was to ensure that the long version of the EQ-HWB should contain at least one item for each of the sub-domains, with more than one item where either this would improve the content validity or items provide information at different levels of a latent construct. The short version is limited by the need to be amenable to valuation using standard methods of preference elicitation with members of the general population.

There were separate rounds of consultation with stakeholders for the long and short versions with the project Advisory Group, Patient and Public Involvement and Engagement (PPIE) group and EuroQol Group membership. To prepare for the consultations, research teams in each of the six countries were asked to summarise and combine the face validity and psychometric performance of each item using a summary score on a 4 to 1 scale (i.e. item performs very well; fairly well; weakly or mixed evidence; or poorly, and why the item is placed in this category e.g. not culturally relevant). Table 1 gives an example of the scoring of item by item for the relationships domain that was presented to consultees in each round.

a) The Long version (EQ-HWB)

The PPIE group was presented with potential items on large, coloured cards, with different colours used for each domain. Members were asked to discuss and allocate items to one of three categories; include, reject or undecided. Participants moved around the room placing the items on large boards for each of the categories. After each domain had been considered, members were given the opportunity to reflect upon their decisions, and they had an opportunity to change the allocation of items. The discussion and final allocation of items were noted.

For the online consultation with the Advisory Group and EuroQoL group members, a majority of the 59 respondents were from the UK (59%). Around half had an academic background, but the second largest group was those involved in resource allocation decision making (29%). Most respondents indicated they wanted to retain all domains in the long version of the measure. However, there was less agreement at the item level when presented with a summary of the evidence on performance of the relevant items for that domain/sub-domain. Table 1 shows the results for the relationships domain. Some items clearly stood out, like 'Loneliness' with 44 consultees wanting to keep it compared to just six looking to drop it (Table 1). The item about 'close to' had 38 indicating it should be dropped. Results were mixed for many items.

These consultation with the PPIE and the stakeholder groups provided a ranking of the preferences of the consultees for the items and this was used by the international research team to select the best items to form the 25-item HWB whilst ensuring there was at least one for each item for each sub-domain (see Table 2). For each sub-domain, the top ranking items were reviewed and one or two items were retained after discussion.

b) Short version (EQ-HWB-S)

Selection of the sub-set of items for the short version reflected two main concerns. Firstly, the focus should be on 'core' generic items relevant for most potential service users. Psychometric evidence around distribution (ceiling and floor effects) and known group validity (by conditions identified across the six countries) was considered to help identify sub-domains relevant to most respondents. Secondly, the correlations between the sub-domains (latent factors) were considered to identify those sub-domains which could be dropped based on their high correlation with remaining sub-domains. Where there was more than one item for a given sub-domain in the long version, we also wanted to ensure we selected the best item.

For this round, the consultation with stakeholders was online with 71 respondents. A majority were from the UK (54%) and identified as academics (65%). Consultees endorsed and ranked items. They were asked to

consider the summary of evidence (e.g., Table 1) and then to indicate whether they strongly recommended, recommended, not sure or do not recommend an item for the short version of EQ-HWB. They were also invited to rank the items and make general comments.

The international team reviewed the results of this consultation in making their decisions about the final set of items for the draft short version that was sent to the Advisory Group for comment. The set of items selected for the draft short version included seven out of the 10 most highly ranked items in the consultation. The main exceptions included the personal needs/self-care item, which was difficult to present in a way that was appropriate for all users including carers, and the item 'I felt sad/depressed' as there was no other item that covered the full severity range of sadness/depression. The term 'depressed' was added even though the standalone item 'I felt depressed' was not recommended after face validity due to concerns over ambiguity in interpreting this as a clinical diagnosis. The combination with sad addressed that concern, whilst also ensuring extremely low mood could be picked up. This resulted in the selection of 9-items covering mobility, daily activities, coping, concentration and thinking clearly, anxiety, sad/depression, loneliness, fatigue and pain

Finally, we considered results from the initial qualitative pilot study of the valuation of a sample of states defined by a draft short version using time trade-off (TTO) and discrete choice experiment (DCE) to inform the selection of items for the short version.[30] A convenience sample of 19 individuals in Sheffield (UK) completed the interview. Prior to the TTO and DCE tasks, respondents were asked to report their own current state using a draft EQ-HWB-S and EQ-5D-5L. Participants valued six states using TTO (three EQ-HWB-S and three EQ-5D-5L) and four EQ-HWB-S states using DCE. After the valuation tasks, they were asked to comment on what they were thinking about when they reported their state using EQ-HWB-S, what was driving their decisions in the TTO and DCE tasks, and to compare undertaking the valuation tasks for EQ-HWB states and EQ-5D-5L states.

The pilot valuation suggested that TTO and DCE are suitable methods to value the draft EQ-HWB-S, though required more time to deliberate compared to EQ-5D-5L.[30] The key finding was that the item on coping was problematic because some respondents made their own judgement on whether they could cope with the state described; some found it inconsistent if it was misaligned to depression, anxiety and exhaustion; and others used it as an overall assessment of all other domains of the state being valued, leading to lexicographic preferences. However, coping was clearly important to respondents. It was decided to replace coping with the control item from the long version since this belongs to the same overall domain and they were highly correlated. The pilot also suggested some simplifications to the response choices to aid presentation of states including from 'some of the time' to 'sometimes'.

The nine items selected from the EQ-HWB for the short version are highlighted in Table 2.

4. DISCUSSION

Overall findings

This paper describes the development of the generic EQ-HWB with two versions, EQ-HWB (25-item) and the EQ-HWB-S (9-items), designed for use in economic evaluation of interventions in health care, social care and public health based on the views of service users and their carers. The development process directly engaged

service users/patients and carers from inception, with input from stakeholders and policy makers who are interested in the ability of a measure to capture the impact of interventions on both health and wellbeing. The EQ-HWB items cover the themes identified from a review of a large body of qualitative evidence on how health and health care, social care and caring roles impact on health and wellbeing. Items were carefully selected on the bases of evidence from face validation and psychometric testing and the views obtained from consultations with stakeholders, resulting in the 25-item EQ-HWB, with a sub-set selected for the shorter nine item EQ-HWB-S**.** Both versions could be used to generate QALYs once a value set is available, since the EQ-HWB-S uses items embedded in the EQ-HWB. The long version provides extra domains that may be of interest in some populations.

The EQ-HWB-S provides an important complement to existing preference-based measures of health like EQ-5D. It covers additional domains like energy ('I felt exhausted') and cognition ('I had trouble concentrating/thinking clearly') that have been recognised as potentially important aspects of health missing from EQ-5D.[7,12] In addition, EQ-HWB-S has domains for social relationships ('I felt lonely') and control found to be important in social care[9] and for caring.[10] As reported in this themed issue, a study comparing EQ-5D-5L and EQ-HWB found significant convergence between the measures for overlapping dimensions.[31] Separating out depression and anxiety increased sensitivity. The EQ-HWB also helped to resolve some ceiling effects and lack of content coverage on the EQ-5D-5L, which suggests it may have greater validity in certain conditions. Whilst these additional domains may not be regarded as important for HTA by agencies who want to focus on health, they are highly relevant for health care more broadly, as well as social care and for assessing the impact on carers.

To ensure the measure is amenable to valuation, dimensions have been excluded that that are only important for some conditions and circumstances including hearing, vision, sleep and feeling safe, acceptance and hope. The longer EQ-HWB version provides a more comprehensive measure that can be used on its own and in future some of these additional items could be examined for valuation. Furthermore, research is being conducted looking at alternative ways to score the longer EQ-HWB using psychometric approaches.

The strength of the EQ-HWB comes from the broadness of the domains covered, the rigour of its development (combining qualitative and quantitative methods) and being international in its development involving six countries covering four continents, something not done in the development of previous preference-based measures. Furthermore, there was a wide range of stakeholders informing the content of the new measure, including patients, service users, members of the public, policy makers, and the scientific community, at each stage of development.

For updates about the EQ-HWB and those interested in using the measure further information is available at

https://euroqol.org/euroqol-is-developing-a-new-instrument-the-eq-hwb-2/.

Limitations and future work

Despite the size and scale of this international project there are important limitations. The review was limited to a targeted selection of reviews of a common medical conditions since it was not possible to look at all conditions and some rarer conditions may be represented less well by this generic measure.

Assessment of known group differences was extensive, but there was limited assessment based on severity within groups, and no assessment of change over time which is an important aspect of the performance of measures; further validation in this context is therefore needed. The groupings used to examine known group validity were very limited for social care users and carers. The psychometric performance requires further testing based on the actual 25 and 9 items versions including studies of equivalence between the direct administration of the EQ-HWB-S and that derived from the same items embedded in the long version. The psychometric analysis included testing in non-English-speaking contexts,[32] but it was not possible to recruit in all groups (e.g., long-term conditions, carers) in all six countries (except for the UK). The results from China indicate further validation will be needed to confirm results from the six countries in this study in other countries and cultures, including Africa where there has been no work to date.

The adoption of a 7-day recall period was a pragmatic decision, and it still may be challenging for conditions that result in less frequent episodes . This will be revisited as more is learned about its performance across different conditions and circumstances. There are other aspects of format and presentation that may be modified with further experience. The impact of any further modifications will need to be tested.

The last stage in the development of the measure is valuation. While piloting results are promising, a full valuation is required using TTO and DCE for further evidence on the feasibility of valuing a large classification system covering health and wellbeing and whether all the items are important in the context of a valuation study.

It could be argued that some existing preference-based measures cover important elements of wellbeing. A comparison of health and social wellbeing (SWB) scales by Richardson and colleagues found that some health utility measures account for a large proportion of variation in SWB, notably AQOL-8D and SF-6D, while EQ-5D accounted for the least.[33] The descriptive systems of AQOL and SF-6D contain some concepts of wellbeing, including affect statements like ‘downhearted and low’ and ‘full of life’, and AQOL also covers self-worth, social isolation, and control. However, SF-6D misses important domains identified in this research. The AQoL-8D is comprehensive, but with 35 items presents considerable challenges for standard valuation methods and was only developed in one country.

Conclusion

The EQ-HWB goes beyond the conventional measures of HRQoL for generating QALYs by encompassing a broader range of generic outcomes including both health and wellbeing. The EQ-HWB is characterized by several important advances: an outcome measure informed by a conceptual framework, designed with extensive input from stakeholders (patients, carers, decision makers involved in HTA), and developed and informed by evidence collected by an international consortium. The EQ-HWB is designed to address the increasing interest in the UK and elsewhere in integrating health and social care, along with outcomes for carers.

REFERENCES:

1. Brazier JE, Ratcliffe J, Tsuchiya A, Solomon J. Measuring and valuing health for economic evaluation. (2nd edition) Oxford: Oxford University Press 2017.

2. Brooks, R., & the EuroQol Group. (1996). EuroQol: The current state of play. *Health Policy, 37*, 53-72.

3. Brazier J, Roberts J, Deverill M. The estimation a preference-based single index measure for health from the SF-36. Journal of Health Economics 2002; 21(2):271-292

4. Feeny DH, Furlong WJ, Torrance GW, Goldsmith CH, Zenglong Z, Depauw S, Denton M, Boyle M. Multiattribute and Single-attribute Utility Function the Health Utility Index Mark 3 system. Medical Care 2002, 40(2):113-128..

5. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics. 2014 Apr;32(4):367-75. doi: 10.1007/s40273-014-0136-z. PMID: 24477679; PMCID: PMC3964297. published in 2010, PharmacoEconomics (online)

6. Herdman, M., Gudex, C., Lloyd, A., Janssen, B., Kind, P., Parkin, D., Bonsel, G. and Badia, X. (2011) Development and preliminary testing of a new five level version of the EQ-5D – the EQ-5D-5L. Quality of Life Research. 20(10), 1727-1736

7. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. The European Journal of Health Economics. 2017 May 30:1-4

8. Mitchell PM, Al-Janabi H, Richardson J, Iezzi A, Coast J. The Relative Impacts of Disease on Health Status and Capability Wellbeing: A Multi-Country Study. PLoS One. 2015; 10(12): e0143590. Published online 2015 Dec 2. doi: 10.1371/journal.pone.0143590 PMCID: PMC4667875

9. Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, Flynn T, Forder J. Outcomes of social care for adults: developing a preference-weighted measure. Health Technology Assessment. 2012;16(16):1-66.

10. Brouwer WB, Van Exel NJ, Van Gorp B, Redekop WK. The CarerQol instrument: a new instrument to measure care-related QoL of informal caregivers for use in economic evaluations. Quality of Life Research. 2006 Aug 1;15(6):1005-21.

11. Rand, S.E., Malley, J.N., Netten, A.P. *et al.* Factor structure and construct validity of the Adult Social Care Outcomes Toolkit for Carers (ASCOT-Carer). *Qual Life Res* **24,** 2601–2614 (2015). https://doi.org/10.1007/s11136-015-1011-x

12. Brazier J, Rowen D, Lloyd A, Karimi M. Future directions in valuing benefits for estimating QALYs: is time up for the EQ-5D? Value in Health 2019, Volume 22, Issue 1: 62-68.

13. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Parkinson J, Secker J, Stewart-Brown S. The Warwick-Edinburgh mental wellbeing scale (WEMWBS): development and UK validation. Health

and Quality of life Outcomes. 2007 Dec;5(1):63.

14. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health related QALY. *Applied Health Economics and Health Policy*. September 2015. Open access: http://link.springer.com/article/10.1007/s40258-015-0194-1?wt_mc=internal.event.1.SEM.ArticleAuthorOnlineFirst.

15. Al-Janabi H, Flynn T, Coast J. (2012) Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Quality of Life Research 21: 167-176 (DOI: 10.1007/s11136-011-9927-2).

16. Coast J, Flynn T, Natarajan L, Sproston K, Lewis J, Louviere J et al. Valuing the ICECAP capability index for older people. Social Science and Medicine 2008; 67(5):874-882.

17. Sen A. Choice, welfare and measurement Harvard University Press, Cambridge, MA (1982)

18. Peasgood T, Carlton J, Brazier JE. A qualitative study of the views of health and social care decision makers on the role of wellbeing in resource allocation decisions in the UK. Economies 2019, 7, 14; doi:10.3390/economies7010014

19. Peasgood, T., Mukuria, C., Carlton, J., Connell, J., Devlin, N., Jones, K., Lovett, R., Naidoo, B., Rand, S., Rejon-Parrilla, J. C., Rowen, D., Tsuchiya, A., **Brazier, J.** What is the best approach to adopt for identifying the domains for a new measure of health, social care and carer-related quality of life to measure quality adjusted life years? Application to the development of the EQ-HWB? *Eur J Health Econ (2021). https://doi.org/10.1007/s10198-021-01306-z*

20. Culyer, A.J. (1989a), Commodities, Characteristics of Commodities, Characteristics of People, Utilities and Quality of Life, in Baldwin, S., Godfrey, C., Propper, C. (eds). The Quality of Life: Perspectives and Policies, London, Routledge

21. Peasgood, T., Mukuria, C., Carlton, J. *et al.* Criteria for item selection for a preference-based measure for use in economic evaluation. *Qual Life Res* **30,** 1425–1432 (2021). https://doi.org/10.1007/s11136-020-02718-9

22. Wilson IB, Cleary PD. Linking clinical variables with HRQoL: a conceptual model of patient outcomes. JAMA. 1995 Jan 4;273(1):59-65.

23. Carlton J, Peasgood T, Mukuria C, Johnson J, Ogden M, Tovey W on behalf of the EQ-HWB PPIE group "Patient and public involvement and engagement (PPIE) within the development of the EQ Health and Wellbeing (EQ-HWB)" HEDS Discussion Paper

24. Mukuria C, Connell J, Carlton J, Peasgood T, Scope A, Clowes M, Rand S, Jones K, Brazier JE "A targeted review of qualitative evidence on domains of quality of life important for patients, social care users and carers to inform the development of the EQ Health and Wellbeing (EQ-HWB)"Value in Health Themed Issue on EQ-HWB

25. Carlton J, Peasgood T, Mukuria C, Connell J, Brazier JE, Ludwig K, Marten O, Kreimeier S, Engel L, Belizan M, Yang Z, Monteiro A, Kuharic M, Luo N, Mulhern B, Greiner W, Pickard AS, Augustovski F "Generation, selection and face validation of items for a new generic measure of quality of life, the EQ Health and Wellbeing (EQ-HWB).". Value in Health Themed Issue on EQ-HWB

26. Bradburn NM, Sudman S, Wansink B. Asking questions: the definitive guide to questionnaire design--for market research, political polls, and social and health questionnaires. John Wiley & Sons; 2004 May 17.

27. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales* (Vol. 1). Oxford: Oxford University Press.

28. Peasgood T, Mukuria C, Brazier JE, Marten O, Kreimeier S, Luo N, Mulhern B, Pickard AS, Augustovski F, Greiner W, Engel L, Yang Z, Gibbons L, Monteiro A, Belizan M, Bjørner J "Developing

a new generic health and wellbeing measure: psychometric survey results for the EQ Health and Wellbeing (EQ-HWB)" Value in Health, Themed Issue on EQ-HWB

29. Li Q, Loke AY. The positive aspects of caregiving for cancer patients: a critical review of the literature and directions for future research. *Psycho-Oncology* 2013, 22(11):2399-2407

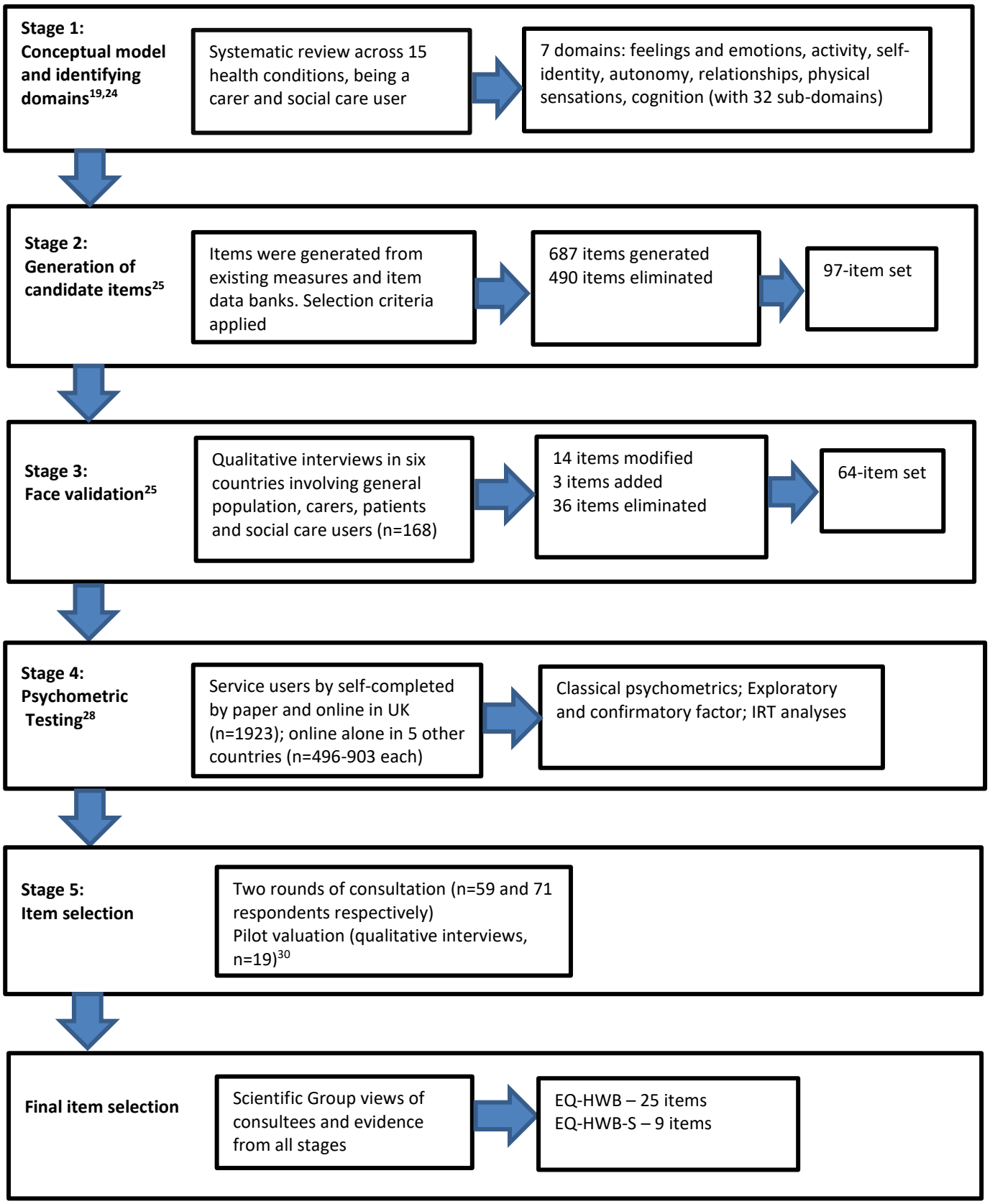
30. Mukuria C, Peasgood T, Brazier. Applying EuroQol Portable Valuation Technology to the EQ Health and Wellbeing Short (EQ-HWB-S): a qualitative pilot study. HEDS Discussion paper

31. Monteiro AL, Kuharic M, Pickard AS. Comparison of the newly developed EQ-HWB and the EQ-5D-5L. Submitted to Value in Health Themed issue on EQ-HWB.

32. Augustovski F, Argento F, Rocío R, Luz G, Mukuria C, Belizán M. The Development of a New International Generic Measure (EQ Health and Wellbeing): Face Validity And Psychometric Stages In Argentina. Submitted to Value in Health Themed issue on EQ-HWB

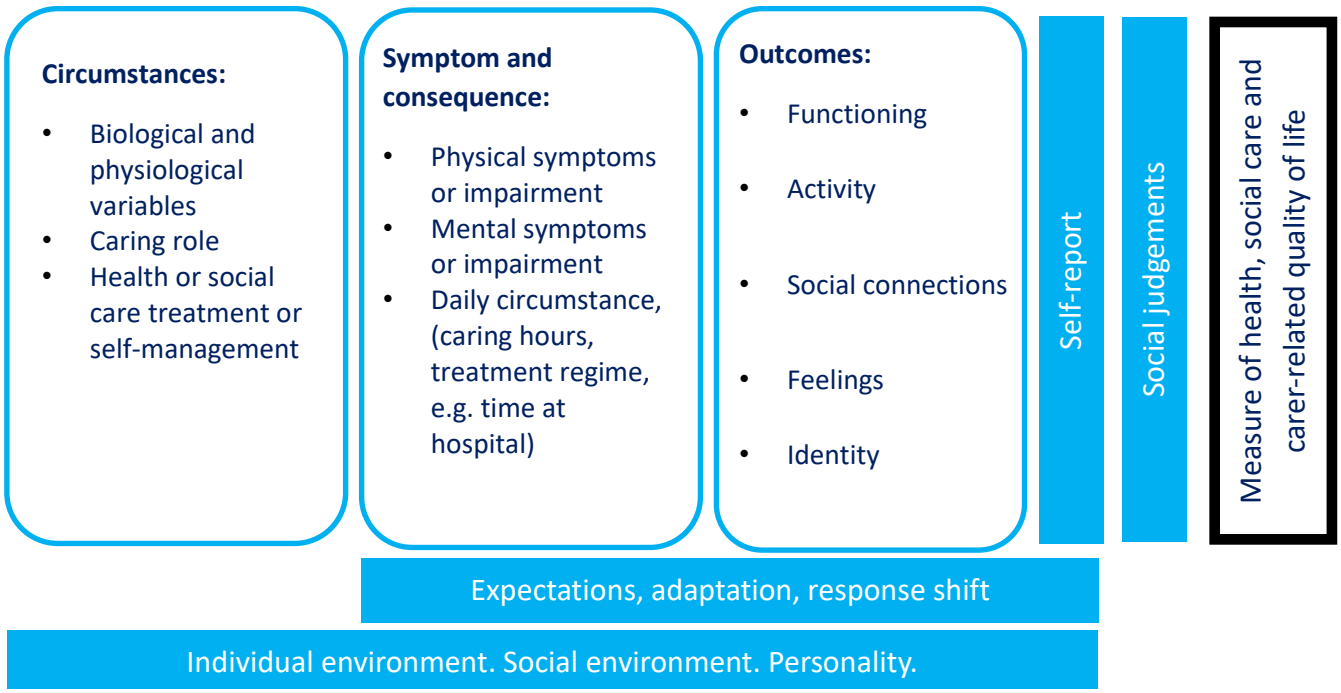
33. Richardson J, Iezzi A, Khan M.A. *et al.* Validity and Reliability of the Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument. *Patient* **7,** 85–96 (2014). https://doi.org/10.1007/s40271-013-0036-x

Figure 1: Development of the EQ-HWB



Note: There were consultations with the PPIE, Advisory and Steering Groups at the end of stages 1,2 and 3, and for final item selection. References to papers describing each stage made in parentheses.

Figure 2: Conceptual model (adapted from Wilson and Cleary)



Note: Whilst there are some dominant causal pathways, there are also many complex bidirectional relationships and interconnections, therefore arrows are not included in the conceptual model.

Figure 3: Preliminary themes and sub-themes from the literature review

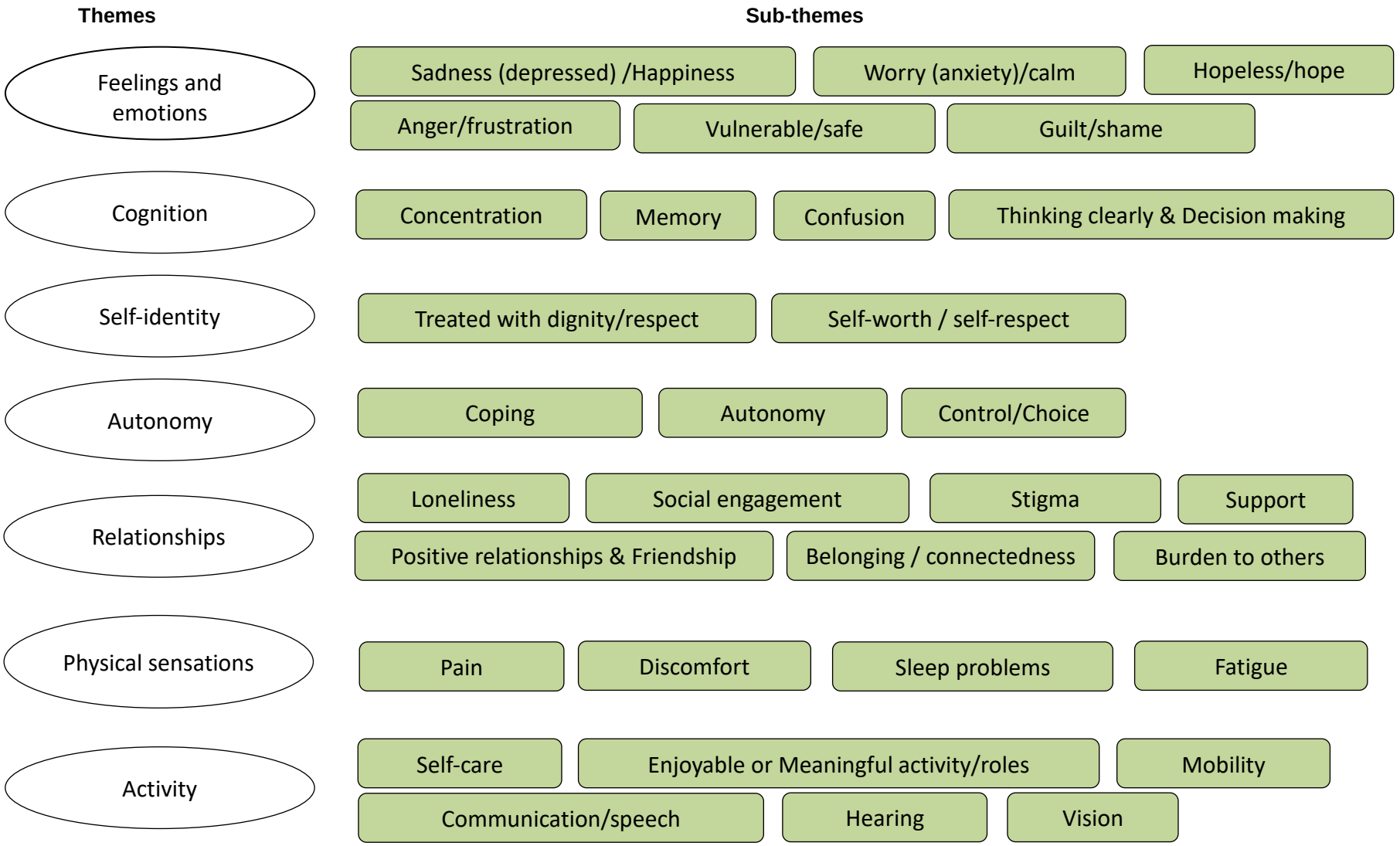


Table 1a: Summary of face validity and psychometric evidence presented in the consultation for the Relationships domain

		UK			Argentina			Australia			China			Germany			USA		
	In the last 7 days…	FV	PV	OV	FV	PV	OV	FV	PV	OV	FV	PV	OV	FV	PV	OV	FV	PV	OV
Support	I felt unsupported by other people (F)	3	4	3	3	3	3	2	3	3	4	3	3	1	3	2	2	3	3
	I had support when I needed it (F)	4	1	1	4	3	2	3	2	2	4	3	3	3	2	3	3	2	2
Positive relations	I got along well with people around me (F)	-	2	2	3	3	3	-	1	1	-	2	3	-	1	1	-	1	1
Lonely	I felt lonely (F)	4	4	4	4	3	3	3	4	4	4	3	4	3	3	3	3	3	3
	I felt there was nobody I was close to (F)	2	4	3	4	3	3	1	3	2	4	4	4	1	3	2	2	4	3
	I felt I had no one to talk to (F)	4	3	3	3	3	3	1	3	2	3	4	4	3	3	3	4	3	3
Stigma	I felt people avoided me (F)	2	3	2	3	2	2	3	2	3	4	3	3	4	3	4	1	3	2
	I felt accepted by others (F)	4	3	3	4	3	3	3	3	3	4	3	3	4	2	3	4	3	3
Belonging	I felt excluded (F)	4	4	4	4	3	3	3	3	3	4	4	4	3	3	3	2	3	2
	I felt left out (F)	4	4	4	4	4	4	3	3	3	4	4	4	4	4	4	4	3	3
	I felt isolated (F)	3	4	3	2	2	2	3	3	3	4	4	4	2	3	2	1	3	2

Notes: F- none of the time, only occasionally, some of the time, often, most or all of the time; FV, PV, OV – face, psychometric and overall validity; Item performs very well (4), fairly well (3), weakly or mixed evidence (2) and poorly (1)

Table 1b. Results of consultation for relationships domain

	Keep in	Drop	Unsure	PI group
Unsupported	18	23	15	drop
Support needed	22	23	11	keep
Lonely	44	6	6	keep
Close to	4	38	14	drop
Talk to	20	22	14	drop

	Keep in	Drop	Unsure	PI group
Avoided	5	38	13	unsure
Accepted	26	15	15	unsure
Excluded	17	24	15	keep
Left out	25	16	15	unsure
Isolated	15	27	14	unsure

Most participants (n=53, 93%) agreed to the inclusion of relationships while 5% (n=3) were unsure and 1 participant said no.

Table 2: EQ-HWB 25 items (items in bold are EQ-HWB-S) - unformatted

These questions are trying to measure how **your** life has been **over the last 7 days**. Please answer all questions. There are no wrong or right answers.

Difficulty (no, slight, some, a lot and unable)

1. How difficult was it for you to see? (*using, for example, glasses or contact lenses if you usually use them*)
2. How difficult was it for you to hear? (*using, for example, hearing aids if you usually use them*)
3. **How difficult was it for you to get around inside and outside*? (using, for example, walking stick, frame or wheelchair, if you usually use them)***
4. **How difficult was it for you to do day-to-day activities? *(for example, working, shopping, housework*)**
5. How difficult was it for you to wash, toilet, get dressed, eat or care for your appearance?

Frequency (None of the time, only occasionally, sometimes, often, most or all the time)

6. I had problems with my sleep
7. **I felt exhausted**
8. **I felt lonely**
9. I felt unsupported by people
10. I had trouble remembering
11. **I had trouble concentrating/thinking clearly**
12. **I felt anxious**
13. I felt unsafe *(fear of falling, abuse or other physical harm)*
14. I felt frustrated
15. **I felt sad or depressed**
16. I felt I had nothing to look forward to

17. **I felt I had no control over my day to day life *(had the choice or do things or have things done for you as you liked and when you wanted)***

18. I felt unable to cope with my day to day life

19. I felt accepted by others *(felt like you were able to be yourself and that you belonged)*

20. I felt good about myself

21. I could do the things I wanted to do

22. I had physical pain

23. I had physical discomfort *(for example, feeling sick, breathless, itching (not including pain))*

Severity (no, mild, moderate, severe, very severe)

24. Physical pain

25. Physical discomfort

© *EuroQol Research Foundation. EQ-HWB™ is a trade mark of the EuroQol Research Foundation*

Questions in **bold** are EQ-HWB-S