Health and Quality of Life Outcomes

RESEARCH Open Access

A framework for extending the health-related quality adjusted life year by combining instruments

Brendan Mulhern[1*], Akanksha Akanksha[1], Richard Norman[2], Mina Bahrampour[1], Peiwen Jiang[1], Deborah Street[1] and Rosalie Viney[1]

Abstract

Current measurement systems focus mostly on health, and not on multiple constructs of quality of life outcomes (for example health and social outcomes) together. This means we don't capture all that is of value to those receiving treatments, and to society more broadly. Recent research has explored how to extend the quality adjusted life year (QALY) beyond a narrow focus on health-related quality of life (HRQoL) aiming to improve the allocation of scarce health and social care resources. Measures of different constructs, including the EuroQol-Health and Wellbeing (EQ-HWB), and different versions of the Adult Social Care Outcomes Toolkit (ASCOT) and ICEpop CAPability (ICECAP), have been developed. Another approach to extending the health focused QALY is to combine existing descriptive systems with different foci into a single instrument. This has the advantage of using available information and allowing trade-offs between the domains of the descriptive systems to be made explicit. The aim of this paper is to propose a framework to guide this approach and outline the methodological process for generating broader descriptive systems. The first section of the paper explains the framework for combining existing instruments and discusses advantages and disadvantages. Advantages include increasing measurement sensitivity to the wider combined quality of life (QoL) impacts of many interventions and using value sets encompassing preferences that are based on trade-offs across diverse constructs. This enables values informed by impacts on broader QoL with relevance across diverse populations, to be used. Disadvantages include theoretical limitations linked to the constructs of QoL included, and practical difficulties combining instruments. The second section of the paper describes the methodological process for generating combined descriptive systems. This includes how to identify which constructs of QoL could be included, and a description of the mixed methods work required to generate a descriptive system that is psychometrically valid, and appropriate for valuation. Combining constructs of QoL from existing instruments offers a promising way to extend the QALY that differs to developing instruments de novo. Future research can use the framework outlined to develop combined instruments and explore the feasibility and wider applicability of the approach, and the use of the instruments generated in resource allocation decision making.

Keywords EQ-5D, Quality of life measurement, Psychometrics, Health state valuation

*Correspondence:
Brendan Mulhern
Brendan.mulhern@uts.edu.au

[1]Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia
[2]Curtin University, Perth, Australia

Introduction

Efficient and equitable allocation of scarce health resources across diverse conditions, interventions and patient populations is fundamental to improving health outcomes for the population. This is often done using cost utility analysis by organisations such as the National Institute for Health and Care Excellence (NICE) in the UK [1], and the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia [2]. As part of cost utility analysis, it is important to measure outcomes that matter to the population when evaluating interventions. However, this is not always the case, with many of the measures used in clinical studies failing to capture important aspects of benefit. Table 1 describes three examples, taken from public summary documents of decisions by the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, that illustrate key benefits of treatments not included in health focused outcome measures. Decision-makers charged with allocating resources face the challenge that many interventions affect not just health (e.g., mobility, pain) but also social outcomes (e.g., autonomy, dignity, social participation, safety). In Australia, the Aged Care Royal Commission has acknowledged that the health of the aging population is integrally linked to these broader aspects of wellbeing, and measuring broader quality of life (QoL) is fundamental to measuring the performance of services [3]. European aged care providers have regularly used social-care QoL measures since 2009 indicating importance of using broader QoL measurement as an indicator for comparing aged care services [4].

Cost utility analysis is informed by the Quality Adjusted Life Year (QALY) Current measurement systems for use in cost utility analysis such as the EQ-5D-5 L [5] or SF-6D [6–9] focus on measuring the construct of health-related quality of life (HRQoL). Constructs are theoretical concepts that are measured empirically and can relate to single or multiple dimensions. HRQoL is a multidimensional construct including that measures the impact of health on quality of life, and includes individual constructs focusing on physical, psychological, and social wellbeing. The EQ-5D and SF-6D systems are examples of generic preference weighted measures (PWMs). These measures have a descriptive system that includes dimensions describing HRQoL that are completed by patients, and a value set which is a set of values estimated from population preferences, and anchored onto the full health – dead utility scale for use in QALY estimation. Recent research has explored extending the conceptualisation of the QALY beyond a narrow focus on HRQoL, arguing that generic scales do not adequately capture the salient dimensions of particular conditions [10], and extending what is measured and valued will improve the allocation of scarce resources [11–12].

A number of initiatives have taken up this challenge and developed instruments that are conceptualised to measure broader constructs, or specifically targeted towards particular populations, de novo. An example of this is the EQ-Health and Wellbeing (EQ-HWB) instrument suite [13–15] which was developed internationally to evaluate interventions in health, public health, and social care. Instruments have also been developed to measure other areas of QoL. This includes the Adult Social Care Outcomes Toolkit-Self Complete Tool(ASCOT-SCT4) [16, 17] which was developed de novo to measure social care related QoL (SCRQoL) of adults with care and support needs, the ICEpop CAPability measure for Adults (ICECAP-A) [18, 19] and older people (ICECAP-O) [20–22], which focus on capabilities, and the Quality of Life – Aged Care Consumers (QoL-ACC) [23–26] specifically for use in aged care.

The instruments above focus on developing measures de novo. Extensive work to develop supplementary additional dimensions (often described as "bolt-ons") to descriptive systems such as EQ-5D has been conducted to extend the core descriptive system to include other HRQoL dimensions [27]. Another area of work develops and adapts broader descriptive systems using existing instruments as the basis. Chen and Olsen [28] argued that, for programme evaluations within wider fields of mental health and community care, the EQ-5D should be complemented by psychosocial bolt-on dimensions, and used the AQoL-8D to test, and in follow up work subsequently develop, bolt-ons for vitality, sleep, personal relationships, and social isolation [29]. Extensive work has developed condition-specific preference-based measures from existing instruments using psychometric and valuation methods [30]. Although many of these condition specific descriptive systems focus on HRQoL, they also extend the measurement framework beyond generic health constructs to include condition specific issues such as cognition (DEMQOL-U for dementia) [31], and symptoms (EORTC QLQ C10D for cancer) [32].

Table 1 PBAC decision summary examples with broader QoL impacts

Submission	PBAC summary
Icatibant for hereditary angioedema	PBAC noted that the benefits of icatibant related to increased *security and control* from the availability of the treatment, rather than from the health gain as a result of the treatment of the attacks.
Poly-L-Lactic Acid for facial lipoatrophy	PBAC noted that the importance of social and psychological impacts was *not captured by the SF-6D*
Tobramycin inhalation powder for cystic fibrosis	PBAC noted the *heavy burden of the standard treatment* and the value of a transportable easy to use device for people with cystic fibrosis, factors that are not captured in health outcome measures.

Taking this further, a novel approach to extending the health focused QALY is to combine descriptive systems measuring different QoL constructs into a broader measurement system incorporating a combined descriptive system with an associated value set estimated from a primary valuation study. The aim of this paper is to outline the framework and methodological process for extending the health-related quality adjusted life year by generating a broader measurement system that combines instruments. This includes outlining the limitations of the approach. This paper provides the conceptual basis for further work to develop and value a broader measurement system using the ASCOT-SCT4 and EQ-5D-5 L instruments as an example.

The paper is divided into three further sections. Section 2 explains the framework for combining existing instruments into broader measurement systems and discusses the advantages and disadvantages of the approach. Section 3 describes the methodological process for generating broader combined measurement systems. Section 4 provides a summary.

Framework for combining descriptive systems

Conceptual definitions – What is being measured and valued and why?

We define a 'broader measurement system' as an approach to combining existing instruments to include both a descriptive system and a value set based on preferences for the dimensions included in the combined instrument. The aim of doing this is to extend the measurement and valuation framework into a broader conceptualisation of QoL. The benefit of doing this is to offer an innovative solution to developing broader measures of QoL to extend the information available for resource allocation decision making based on existing instruments.

The combined descriptive system measures different broad constructs of QoL using the dimensions included in the source instruments The theoretical utility that is measured is either a latent (unanchored) utility, or a utility anchored onto the full health – dead scale that combines preferences for the different generic, condition or population specific QoL constructs combined as a single value set (as it is estimated from trade-offs between dimensions measuring diverse QoL constructs rather than using value sets derived from instruments including dimensions measuring narrower constructs of QoL (e.g. HRQoL).

There are many preference weighted instruments that have a narrow focus on a particular construct of QoL, and the framework suggested here is applicable across instruments. However, a strong theoretical and empirical basis is required to establish where broad frameworks are needed. Different QoL constructs that could be conceptualised to be complementary, and relevant to the assessment of interventions for particular populations, are key candidates for this approach. Another area where this approach shows promise is in the combination of generic and condition specific constructs, which results in a more holistic and in-depth approach to measuring the impacts of a condition, and allows for decision making to be informed by trade-offs across diverse generic and condition specific domains. This also potentially extends the use of generic measures such as EQ-5D into condition specific frameworks using value sets informed by trade-offs across broader domains.

What are the advantages and disadvantages of this approach?

There are theoretical and practical advantages and disadvantages to a broad measurement system combining QoL constructs. The approach provides a unique solution by using valuation methods to combine two widely used existing instruments that between them measure distinct outcomes into a single value set, so that they can be used together to compare the outcomes of interventions using a broader perspective (but collected in their original format). Therefore, advantages include increasing measurement sensitivity (i.e. being sensitive to differences between groups, and to change over time) to the wider QoL impacts of many interventions by administering a more comprehensive descriptive system. The broader descriptive system can benefit from the already available evidence regarding the measurement characteristics of, and relationship between, the instruments included. The value sets developed for the combined system are based on preferences that are a result of trading off between diverse QoL constructs. This enables values informed by impacts on broader QoL concepts with relevance across diverse health conditions, populations, and interventions, to be used in decision making. Another benefit for sensitivity analysis is that the results using value sets for the individual instruments can be compared with those generated using the combined version. Finally, the combined value set can be applied to any existing data that includes both instruments. Mapping methods can also develop approaches to estimate values for the combined instrument when only one of the single instruments is collected.

There are a range of challenges related to the approach that need to be acknowledged. These include theoretical limitations linked to the constructs of generic or condition specific QoL the extended instrument could include (therefore still potentially limiting the coverage of the combined system in certain conditions or populations). As many constructs of QoL from different instruments overlap, there is the potential for double counting. This can be assessed during the development phase using established psychometric approaches such

as to assess item redundancy (i.e. items measuring the same underlying construct). Redundant items can be removed, however changing or reducing the items on existing instruments can be challenging practically and requires input and support from the original instrument developers.

There are also difficulties in combining instruments into a combined system that is amenable to valuation, as combining items from different instruments with different item formats generates complex descriptive systems for valuation. Complexities include different item descriptors and response scales, a large number of dimensions for valuation, and the potential for both positive and negatively worded items, which increase the cognitive challenge of completing valuation tasks.

The example of HRQoL and SCRQoL

Using an example supports conceptualisation of this approach. Consider HRQoL and SCRQoL, which have been selected here as there is likely to be significant benefit for policy makers from assessing the outcomes of health interventions that impact both health and social care. The instruments selected to measure HRQoL and SCRQoL are the EQ-5D-5 L [5] and the ASCOT-SCT4 [17] respectively.

The EQ-5D-5 L measures HRQoL across five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) with five response levels (none, slight, moderate, severe, extreme/unable to). More than 25 value sets have been developed for use in resource allocation decision making [33]. The ASCOT-SCT4 measures SCRQoL across eight dimensions (control, cleanliness and comfort, food and drink, safety; social participation and involvement; occupation; accommodation cleanliness, dignity) and four response levels using a variety of wordings reflecting level of need. A number of value sets are available around the world.

The EQ-5D-5 L and ASCOT-SCT4 are strong candidate measures to assess the constructs as they are extensively validated instruments with existing evidence regarding their measurement relationship (34–35) and have widespread use in policy contexts. The utility measured by an instrument combining EQ-5D-5 L and ASCOT-SCT4 is a scale assessing health related and SCRQoL. The instrument would enable a more diverse and patient relevant range of outcomes of health care to be explicitly measured and valued for use in economic evaluation and priority setting, including aspects of social wellbeing such as autonomy, dignity, and social participation, alongside measures of health such as mobility and pain. This is important, as internationally there is a rapidly aging population that will require increased funding for health and social care [36].

Methodological process to develop combined measurement system

In this section of the paper, we describe the methodological process for generating combined descriptive systems and raise discussion points. The methods we propose are outlined in Fig. 1 and provide a five stage generalisable approach to combining and translating data from different measures of outcome into a combined measurement system that is valued on a single scale. Each of the stages is described in detail, with reference to the pilot work conducted by the authors, below.

Stage 1 - Establish conceptual basis for combining constructs and identifying instruments

It is important to first establish the basis for generating combined measurement systems. This could be established by developing a measurement model, which is a model that describes the areas of QoL required to measure an overall multidimensional construct, and the relationships between them [37]. This supports the identification of instruments that align with the broader constructs measured.

Alternatively, there are a number of empirical approaches to identifying where existing instruments display limitations, both conceptually, for specific populations, and in comparison to other instruments. These methods also have benefits in establishing where bolt-ons to the EQ-5D may be required (see Sect. 2.2). For example, qualitative content validity assessment (which assesses the extent to which instruments measure all aspects of a construct) can be used with patient groups to identify the domains of importance that may not be covered by single instruments [35, 38, 39]. Psychometric methods such as factor analysis have also been used to explore how instruments measuring diverse constructs complement and extend the measurement of QoL (40–41). Exploring evidence from these approaches using primary and secondary analyses, and reviews of the literature, can be used to support the need for a broader measurement system, and identify and test psychometrically validated instruments.

The combination of HRQoL and SCRQoL used in the authors' example conceptualises EQ-5D and ASCOT-SCT4 as instruments with little overlap in the constructs measured, that would therefore provide complementary information across the dimensions included, and values used to inform decision making. This is because HRQoL is likely to be affected by the health conditions that populations such as those in aged care, palliative care, and disability care experience, the SCRQoL impacts of interventions will not be specifically measured, except to the extent that they are indirectly reflected in changes in HRQoL. Consequently, SCRQoL impacts will not be reflected in the values used for decision-making.

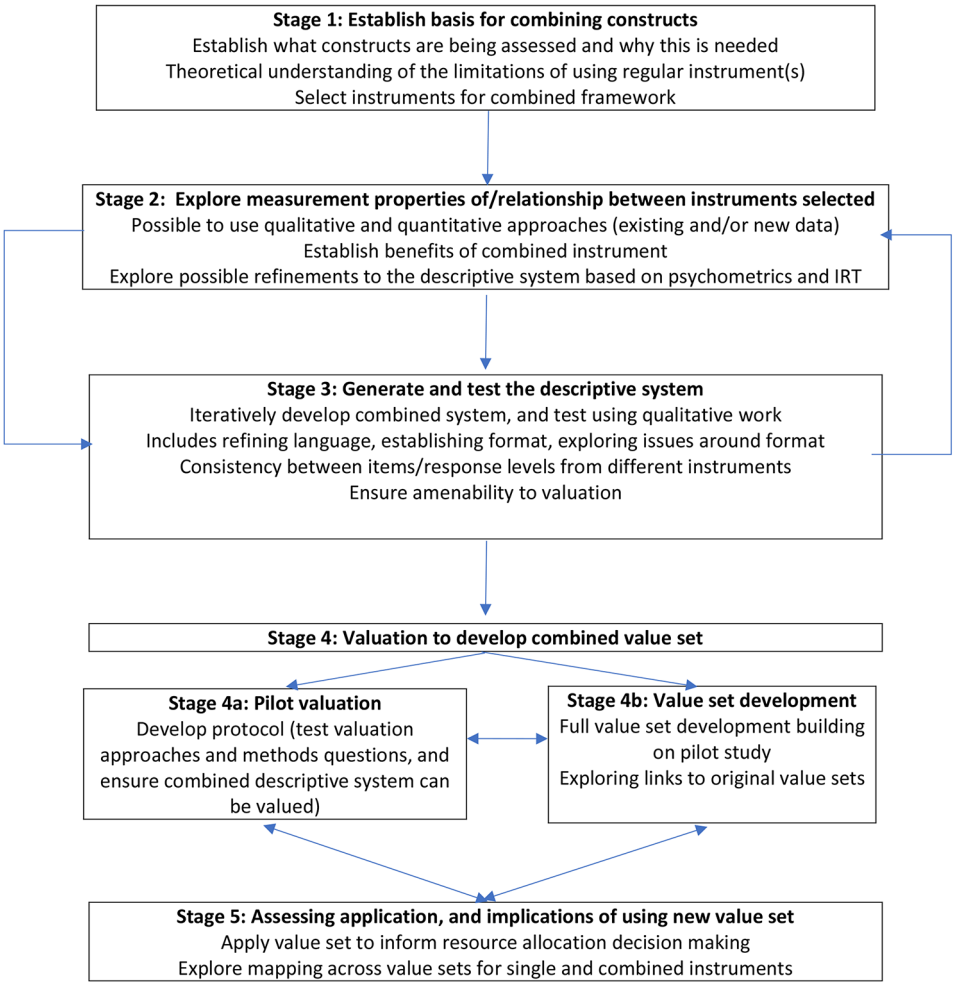

Fig. 1 Methodological process for developing a combined measurement system

Therefore, a combined system with values reflecting both areas of QoL on the same scale is advantageous.

Stage 2 - Explore measurement properties of, and relationship between, the instruments selected

A key stage of the development of the combined descriptive system is to explore measurement properties of the instruments. The measurement properties include characteristics of instrument such as content and construct validity, and responsiveness to change over time. Systematic investigation is required to understand the performance of the measures because it is important to understand what the instruments are measuring, and also the relationship between diverse instruments. This stage is part of an iterative process alongside Stage 3 (generating the descriptive system). It builds on the evidence from existing literature about the measurement properties of the instruments, and has the potential to extend the evidence base tackle any limitations found with the earlier work, and triangulate the evidence to inform the development of a combined system. For example, comparative psychometric work is sample specific, and therefore may not be applicable to the country in which the measure is being developed, or the majority of the target population. Also, the existing evidence may be limited in the analysis methods used, and further work may be required to explore other analytical approaches.

During Stage 2, psychometric analysis is conducted to assess the relationship between the dimensions, and the dimension structure, to understand in which settings, conditions and populations different instruments should be used, and also understand the relationship between them. It is important to establish where there is convergence and divergence in what is measured by the domains across each instrument. Methods employed include classical psychometric approaches [42] to assess construct validity, regression to model the extent to which domains on one measure explain outcomes on the other, and assessment of the dimension structure using factor analytic approaches.

Stage 3 - Generate and test the descriptive system

The generation of the combined descriptive system is informed by a range of practical issues as well as the empirical work to understand the relationship between instruments, and has a number of requirements. The first requirement is that the combined measurement system must have valid measurement properties, and avoid dimension or item redundancy. The second requirement is that the measurement system must be amenable to valuation and the generation of a combined value set. This includes considering the wording of the dimension levels, and potential simplifications to ensure interpretability whilst retaining the meaning of the original item, and the number of dimensions included in the descriptive system. Both requirements are considered in the formation of the system at Stage 3, with the descriptive system refined iteratively.

Regarding the first requirement, the results from the Stage 2 psychometric work, and evidence from the existing literature, can be used to identify poorly performing dimensions, or dimensions that directly overlap, meaning that one is redundant. We then need to consider removing dimensions from the descriptive system valued. This means that the instruments are no longer presented as intended by the developers, and the original value sets cannot be estimated, but it may result in a system with stronger measurement properties.

In the example of combining HRQoL and SCRQoL, previous work has found that there is limited convergence between the dimensions included [35], and more recent empirical work in large samples of the Australian population with a range of common health conditions has supported this. In this work, the EQ-5D-5 L dimension that converged most with the ASCOT-SCT4 was Anxiety/depression, but correlations were low to moderate [43]. This evidence can be triangulated to support the generation of a combined measurement system including all 13 dimensions.

Regarding the second requirement, the relationship between the dimensions and the wording needs to be examined in a systematic way. Ensuring the number of dimensions included is amenable to valuation is also an important consideration. Qualitative evidence can inform the development of the descriptive system, including cognitive interviews to assess the length and wording of the descriptive system, and compare different possible framings of the dimensions as overall items and in valuation task settings. Using consistent descriptions and severity levels across the constructs could lead to further insight regarding the relationship between the different dimensions. It may also lead to the development of dimension descriptors that are quite different to those in the original instruments, but further work could test whether homogenising the wording influences both self-report and trading across dimensions. The developmental work conducted for the instruments included is used as a key starting point in developing and refining a descriptive system, and understanding how respondents trade across dimensions. Any refinements would be assessed for consistency with the original instruments in terms of both response patterns and valuation responses. Practically, any changes would need to be discussed with the original developers and approved in line with their intellectual property and adaptation policies.

The practical challenges of combining instruments into a descriptive system amenable for valuation are demonstrated by the EQ-5D-5 L and ASCOT-SCT4 First, valuing 13 dimensions may raise challenges for respondents. Second, the wording and length of the dimensions and the number and consistency of the response levels differs between the EQ-5D-5 L and ASCOT-SCT4. This has potential challenges for the interpretation of the dimensions, and conceptualisation of the overall QoL construct that is described. Although there are challenges, it should be noted that pilot work using online DCE found relatively interpretable and consistent estimates [44].

Stage 4 - Valuation to develop combined value set

In this section, methods to develop combined value sets, and considerations in this process are outlined. In this example, we focus on the use of online DCE methods for the purpose of valuation. We argue that DCE is the valuation method best placed to explore the ideas outlined in this paper. This is because DCE has the flexibility to test a range of methodological and design questions by presenting different versions in large samples relatively cheaply and efficiently. It has become widely established in recent years [45–46], in particular the development and use of DCE with duration [47–49] which can allow the estimates to be anchored onto the full health to dead utility scale measuring the combined constructs of QoL included in the descriptive system. However, we acknowledge that other valuation approaches, including different DCE implementations, could be used. As this framework likely to produce relatively long descriptive systems for valuation, which can be challenging for respondents. Innovation in the design and presentation of valuation studies, building on previous work aiming to simplify the valuation process such as imposing dimension level overlap [50], testing presentation methods [51] and presenting fewer attributes to respondents whilst still valuing the overall descriptive system [52], should be explored.

Pilot valuation work

We propose that pilot work is conducted to test methodological and descriptive system issues prior to embarking on a full valuation, and test the DCE valuation approach proposed. However, we recognise that this is not always

possible. Nevertheless, if it has been demonstrated that descriptive systems from preference weighted instruments measuring different constructs can be combined, and a draft descriptive system for valuation has been constructed, it is important to understand whether they can be valued on the same scale (and inform the protocol for the main value set data collection).

It is also important to qualitatively assess the DCE tasks using cognitive interview methods. A pilot study can address questions relating to the amenability of the combined descriptive system for analysis, the difficulty of the valuation tasks for respondents, and approaches to simplify this as outlined above. It can also be used to explore possible preference heterogeneity. This is important, as it could be that there are cross-cultural differences in preferences for different aspects of QoL depending on factors such as the characteristics of the healthcare system, and different attitudes towards different QoL constructs.

Value set development

The development of the value set for the combined descriptive system is modelled from data collected according to a protocol informed by the pilot work.

Stage 5 - Assessing application and implications of combined value set

It is important to explore the practical decision-making implications of using a value set for a combined QoL instrument, such as the one described here. Assuming interventions that improve a more health-focused measure of QoL (such as the EQ-5D-5 L) also have a positive impact on broader QoL and wellbeing, then the incremental QALYs generated for interventions using a combined instrument will be greater, and the ICERs will be lower. While we believe this to be the most likely consequence, we also expect there to be considerable heterogeneity around these effects; indeed, if there was a direct transformation of QALYs generated using the EQ-5D to QALYs generated using a combined instrument, then the benefit of the combined instrument would be diminished as QALY maximisation would lead to the same decisions being made. The expected reduction in ICERs from using a combined instrument does pose a challenge in settings with well-established rules of thumb around acceptable ICER thresholds. Therefore, we recommend exploration of the relationship between QALYs generated using single instruments, QALYs estimated from mapping the single instrument to the combined value set, and those estimated using the value set from the combined instrument, to help inform policy makers.

Discussion

This paper outlines a framework for combining constructs of QoL from existing instruments and offers a way to extend the QALY that differs to developing instruments de novo. We argue that this approach provides an innovative solution to the problem of narrow focus on single constructs of QoL by allowing for the development of new and original tools for measuring and valuing health that combines outcomes and values them on the same scale.

There are a number of key benefits of this approach that may support the use of combined descriptive systems in a number of settings. The methods used provide scope to expand preference-based indices to produce values that are more sensitive to the impacts of different interventions and services across a wider range of patient groups and settings. This means that the combined descriptive systems and associated value sets be beneficial for use in Health Technology Assessment if an intervention has broader impacts, and if broader measures have the potential to be accepted as part of HTA guidelines. Studies can be designed to include the relevant instruments, but another benefit, and potential use, is the ability to apply the combined value set to assess existing datasets including the measures. This is an advantage over measures developed de novo, and means that multiple value sets from the individual and combined measures can be applied to the same datasets in sensitivity analyses, and comparisons can also be made within and across studies.

In this paper, the example used was a descriptive system combining HRQoL and SCRQoL. However, the framework is not limited to this, and other combinations may be considered where there is a theoretical basis, including a justifiable target sample, or intervention with impacts on multiple constructs that a combined measure would be sensitive to, in comparison to a single instrument. For example, the preference relationship between HRQoL and wellbeing may be of interest across a range of sectors. Another avenue for exploration would be in developing a combined measure targeting both generic and condition specific constructs. For example, combining generic dimensions with condition specific instruments or PWMs assessing symptoms and side effects for use in the assessment of cancer treatments. The same framework developed in this study could be used to explore the development of these instruments, and valuation work could explore trade-offs between generic and condition specific dimensions of QoL to assess preferences in general and patient populations. Although there are multiple use cases in primary and secondary studies for a combined measure, several issues need to be considered. It is important to consider burden on patients given the potential addition of items. Further, the relevance of the combined measure for the study population

should be carefully considered it light of whether a single measure could be used. A number of other key limitations in developing and using combined descriptive systems should be acknowledged. Theoretically, it may not be conceptually relevant to combine multidimensional QoL constructs, for example if the broader combined measure may have limited applicability to different populations and interventions, or if there is substantial overlap between the constructs. It may also not be valid to combine multidimensional generic and condition specific constructs if the condition substantial overlap with generic measures, or alternatively is not expected to impact of generic HRQoL. There are also practical and use limitations. As noted above, combining instruments raises practical challenges for valuation, and solutions to these (e.g. modifying the dimension wording) needs to ensure that the intended meaning of the original item is maintained, as this ensure that the values can be applied validly to existing datasets.

In summary we have developed a framework and methodological process for combining descriptive systems with different foci into the same measurement and valuation framework, and demonstrated how this could be operationalised using HRQoL and SCRQoL as example constructs. This framework provides the basis for the further development of combined descriptive systems, testing the limitations of the approach, and exploring the use of these instruments in primary and secondary studies to facilitate the accurate measurement of outcomes and inform resource allocation.

Abbreviations

ASCOT-SCT4	Adult Social Care Outcomes Toolkit – Self Complete Tool
EQ-HWB	EuroQol – Health and Wellbeing
HRQoL	Health Related Quality of Life
ICECAP	ICEpop CAPability measure
PBAC	Pharmaceutical Benefits Advisory Committee
QALY	Quality Adjusted Life Year
QoL	Quality of Life
SCRQoL	Social Care Related Quality of Life

Acknowledgements
We thank the project advisory team (Richard De Abreu Lourenco, Lidia Engel, Louise Longworth, Juliette Malley, David Whitehurst) for their expert advice and guidance, and participants in the EuroQol 2023 Plenary meeting for their helpful comments on an earlier draft.

Author contributions
All authors made substantial contributions to the conception, drafting and revision of this work, and have approved the submitted version.

Funding
This work is funded by an Australian Research Council Discovery Project (DP210102021) and the EuroQol Research Foundation (RA-148).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval
Not applicable.

Competing interests
BM, RN, RV and DS are members of the EuroQol Research Foundation.

Received: 31 July 2024 / Accepted: 27 February 2025
Published online: 24 March 2025

References

1. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. London: NICE; 2023.
2. Pharmaceutical Benefits Advisory Committee. Guidelines for Preparing submissions to the pharmaceutical benefits advisory committee. Canberra: Australian Department of Health; 2015.
3. Royal Commission into Aged Care Safety and Quality. Final Report: Care, Dignity and Respect. Canberra, 2021.
4. Siette J, Jorgensen ML, Georgiou A, et al. Quality of life measurement in community-based aged care – understanding variation between clients and between care service providers. BMC Geriatr. 2021;21:390.
5. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.
6. Brazier JE, Roberts J, Deverill M. The Estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271–92.
7. Brazier JE, Roberts J. The Estimation of a preference-based measure of health from the SF-12. Med Care. 2004;42(9):851–9.
8. Brazier J, Mulhern B, Bjorner JB, Gandek B, Rowen D, Alonso J, Vilagut G, Ware J. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. Med Care. 2020;58(6):557–65.
9. Mulhern B, Bansback N, Norman R, Brazier J, Valuing. SF-6Dv2 in the UK using a discrete choice experiment with duration. Med Care. 2020;58(6):566–73.
10. Neumann PJ, Cohen JT. QALYs in 2018—Advantages and concerns. JAMA. 2018;319(24):2473–74.
11. Brazier J, Tsuchiya A. Improving Cross-Sector comparisons: going beyond the Health-Related QALY. Appl Health Econ Health Policy. 2015;13(6):557–65.
12. Brazier JE, Rowen D, Lloyd A, Karimi M. Future directions in valuing benefits for estimating QALYs: is time up for the EQ-5D? Value Health. 2019;22(1):62–8.
13. Brazier J, Peasgood T, Mukuria C, Marten O, Kreimeier S, Luo N, Mulhern B, Pickard AS, et al. The EQ health and wellbeing: overview of the development of a measure of health and wellbeing and key results. Value Health. 2022;25(4):482–91.
14. Carlton J, Peasgood T, Mukuria C, et al. Generation, selection, and face validation of items for a new generic measure of quality of life: the EQ health and wellbeing. Value Health. 2022;25(4):512–24.
15. Peasgood T, Mukuria C, Brazier J, Marten O, Kreimeier S, Luo N, Mulhern B, Greiner W, Pickard AS, Augustovski F, Engel L, Bjorner J, et al. Developing a new generic health and wellbeing measure: psychometric survey results for the EQ health and wellbeing (EQ-HWB). Value Health. 2022;25(4):525–33.
16. Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, Flynn T, Forder J, Wall B. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol Assess. 2012;16(16):1–166.
17. Adult Social Care Outcomes Toolkit. https://www.pssru.ac.uk/ascot/. Accessed 16/01/2025.
18. Al-Janabi H, Flynn T, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res. 2012;21(1):167–76.
19. Flynn T, Huynh E, Peters TJ, Al-Janabi H, Clemens S, Moody A, Coast J. Scoring the Icecap-a capability instrument. Estimation of a UK general population tariff. Health Econ. 2015;24(3):258–69.
20. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? Soc Sci Med. 2006;62:1891–901.
21. Coast J, Flynn T, Natarajan L, Sproston K, Lewis J, Louviere J, et al. Valuing the ICECAP capability index for older people. Soc Sci Med. 2008;67(5):874–82.
22. ICEpop CAPability Measures. https://www.bristol.ac.uk/population-health-sciences/projects/icecap/. Accessed 16/01/2025.

23. Cleland J, Hutchinson C, McBain C, et al. Developing dimensions for a new preference-based quality of life instrument for older people receiving aged care services in the community. Qual Life Res. 2021;30(2):555–65.
24. Ratcliffe J, Bourke S, Li J, Mulhern B, Hutchinson C, Khadka J, Milte R, Lancsar E. Valuing the quality of life for aged care consumers (QoL-ACC) instrument for quality assessment and economic evaluation. PharmacoEconomics. 2022;40(11):1069–79.
25. Hutchinson C, Ratcliffe J, Cleland J, et al. The integration of mixed methods data to develop the quality of life—aged care consumers (QOL-ACC) instrument. BMC Geriatr. 2021;21:702.
26. Quality of Life-Aged Care Consumers. https://www.qol-acc.org/quality-of-life-instrument. Accessed 16/01/2025.
27. Geraerds AJLM, Bonsel GJ, Janssen MF, et al. Methods used to identify, test, and assess impact on preferences of Bolt-Ons: A systematic review. Value Health. 2021;24(6):901–16.
28. Chen G, Olsen JA. Extending the EQ-5D: the case for a complementary set of 4 psycho-social dimensions. Qual Life Res. 2023;32:495–505.
29. Chen G, Olsen JA. Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. Qual Life Res. 2020;29(11):3119–29.
30. Rowen D, Brazier J, Ara R, Zouraq IA. The role of Condition-Specific Preference-Based measures in health technology assessment. PharmacoEconomics. 2017;35(S1):33–41.
31. Mulhern B, Rowen D, Brazier J, Smith S, Romeo R, Tait R, Watchurst C, Chua K-C, Loftus V, Young T, Lamping D, Knapp M, Howard R, Banerjee S. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technol Assess. 2013;17(5):1–140.
32. King MT, Viney R, Pickard AS, et al. Australian utility weights for the EORTC QLU-C10D, a Multi-Attribute utility instrument derived from the Cancer-Specific quality of life questionnaire, EORTC QLQ-C30. PharmacoEconomics. 2018;36(2):225–38.
33. Devlin N, Roudijk B, Ludwig K. Value sets for EQ-5D-5L: A compendium. Comparative Review & User Guide. Netherlands: Springer; 2022.
34. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021;30(3):647–73.
35. Van Leeuwen KM, Jansen APD, Muntinga ME, et al. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. BMC Health Serv Res. 2015;15:201.
36. Appleby J. Spending on health and social care over the next 50 years: why think long term? London: The King's Fund; 2013.
37. Atkinson MJ, Lennox RD. Extending basic principles of measurement models to the design and validation of patient reported outcomes. Health Qual Life Outcomes. 2006;4:65.
38. Matza L, Boye K, Stewart K, et al. A qualitative examination of the content validity of the EQ-5D-5L in patients with type 2 diabetes. Health Qual Life Outcomes. 2015;13:192.
39. Keeley T, Al-Janabi H, Lorgelly P, Coast J. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. PLoS ONE. 2013;8(12):e85287.
40. Finch A, Brazier JE, Mukuria C, Bjorner JB. An exploratory study on using Principal-Component analysis and confirmatory factor analysis to identify Bolt-On dimensions: the EQ-5D case study. Value Health. 2017;20(10):1362–75.
41. Finch A, Mulhern B. Where do measures of health, social care and wellbeing fit within a wider measurement framework? Implications for the measurement of quality of life and the identification of bolt-ons. Soc Sci Med. 2022; 313.
42. Fayers P, Machin D. Quality of life: the assessment, analysis and interpretation of Patient-Reported outcomes. London: Wiley.
43. Akanksha A, Mulhern B, Street D, Viney R. Psychometric insights into the EQ-5D-5L and EQ-HWB in a sample of older Australians. Netherlands: EuroQol Plenary. Noordwijk; 2024.
44. Mulhern B, Norman R, DeAbreu Lourenco R, Street D, Malley J, Viney R. Investigating the relative value of health and social care related quality of life using a discrete choice experiment. Soc Sci Med. 2019;233:28–37. https://doi.org/10.1007/s40273-018-0714-6.
45. Mulhern B, Norman R, Street D, Viney R, One Method. Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. Pharmacoeconomics; 2019; 37(1): 29–43.
46. Wang H, Rowen DL, Brazier JE, Jiang L. Discrete choice experiments in health state valuation: A systematic review of progress and new trends. Appl Health Econ Health Policy. 2023;21(3):405–18.
47. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ, 2012: 306–18.
48. Norman R, Mulhern B, Lancsar E, Lorgelly P, Ratcliffe J, Street D, Viney R. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. PharmacoEconomics. 2023;41(4):427–38.
49. Roudijk B, Jonker M, Bailey H, Pullenayegum E. A direct comparison between discrete choice with duration and composite time Trade-Off methods: do they produce similar results?? Value Health. 2024;27(9):1280–8.
50. Jonker M, Donker B, De Bekker-Grob E, Stolk E. Effect of level overlap and color coding on attribute Non-Attendance in discrete choice experiments. Value Health. 2018;21(7):767–71.
51. Mulhern B, Norman R, Shah K, Bansback N, Longworth L, Viney R. How should discrete choice experiments with duration choice sets be presented for the valuation of health States? Med Decis Making. 2018;38(3):206–318.
52. Sullivan T, McCarty G, Ombler F, Turner R, Mulhern B, Hansen P. Creating an SF-6Dv2 social value set for new Zealand. Soc Sci Med. 2024; 345.

Publisher's note