The European Journal of Health Economics
https://doi.org/10.1007/s10198-026-02009-z

RESEARCH

A consistency criterion for EQ-5D-5L bolt-on valuation

Michał Jakubczyk[1] · **Fanni Rencz**[2,3] · **Kim Rand**[4,5]

Received: 22 April 2026 / Accepted: 12 October 2026
© The Author(s) 2026

Abstract

Objective EuroQol is developing the EQ-5D Bolt-on Toolbox, an experimental set of bolt-ons designed to improve the sensitivity of EQ-5D-5L in specific populations. For use in cost-utility analyses, EQ-5D-5L+bolt-on value sets should be consistent with EQ-5D-5L value sets. This study proposes a method to define and maintain such consistency and examines its properties.

Methods We developed a novel *contraction method* that defines consistency using the distribution of disutilities in measurement data and estimates a contraction parameter from the conditional distribution of bolt-on levels given the core EQ-5D-5L profile to account for information overlap. The method was tested for nine bolt-ons: breathing, cognition, hearing, self-confidence, skin irritation, sleep, social relationships, tiredness, and vision. We used general-population samples from 14 countries, with 10,395–42,206 observations per bolt-on, and patient samples from six disease areas, with 312–2,486 observations. Core-dimension disutilities were based on the UK EQ-5D-5L value set, and bolt-on disutilities were taken from published studies. Associations of the contraction parameter with bolt-on disutilities and core-dimension correlations were assessed using linear regression.

Results Achieving consistency required contraction factors ranging from 0.871 for breathing to 0.989 for sleep. Stronger contraction was required for bolt-ons with larger disutilities and for those more strongly correlated with the core dimensions, although only the former association was statistically significant.

Conclusion The contraction method aligns EQ-5D-5L+bolt-on value sets with EQ-5D-5L value sets, reducing potential double counting while preserving the increased sensitivity provided by bolt-ons.

Keywords EQ-5D-5L · Bolt-on · Valuation · Consistency · EQ-5D Bolt-on Toolbox

JEL classification I10 · D61 · C18 · D63

Introduction

To inform reimbursement decisions, costs and effects of health technologies are compared in cost-utility analysis (CUA). Health technologies offer various health benefits: prolonging longevity or improving various aspects of health related quality of life (HRQoL). Because different technologies compete for resources from a single budget, health benefit measures should be comparable across diseases. With that goal, usually quality-adjusted life years (QALYs) are used (Whitehead and Ali 2010).

QALYs are calculated based on the expected survival time and the numerical value representing HRQoL. To improve comparability across diseases, generic instruments are used to measure HRQoL, with the EQ-5D family of instruments being the most widely used (Kennedy-Martin et al. 2020). For these instruments, HRQoL is assessed using five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD), henceforth *the core* dimensions (CDs). In EQ-5D-5L, the focus of this study, each dimension is measured using five levels:

✉ Michał Jakubczyk
michal.jakubczyk@sgh.waw.pl

1 SGH Warsaw School of Economics, Warszawa, Poland

2 Department of Health Policy, Corvinus University of Budapest, Budapest, Hungary

3 EuroQol Research Foundation, Rotterdam, The Netherlands

4 Akershus University Hospital, Lørenskog, Norway

5 Maths in Health B.V., Klimmen, The Netherlands

Published online: 29 October 2026

1 denoting no problems, and 5 denoting extreme problems/unable to. Hence, a health state is represented by a string of five digits, and 11111 represents no health problems, while 55555 represents the worst EQ-5D-5L health state (the *pits state*). Using valuation studies, EQ-5D-5L health states are assigned values — *QALY weights*, *index weights*, or *utilities* — expressing how good or bad these states are. These weights are collected in *value sets* (Golicki et al. 2019; Pickard et al. 2019; Versteegh et al. 2016).

The wide use of EQ-5D-5L in CUA is due to its simplicity, sensitivity, and ability to capture HRQoL in multiple health conditions (Feng et al. 2021). However, for some conditions, its sensitivity can be reduced (Payakachat et al. 2015), for example, through the *ceiling problem*, i.e., no problems being reported in CDs despite the presence of disease-specific symptoms. While extending the number of levels from 3 in EQ-5D-3L to 5 in EQ-5D-5L increased the sensitivity (Buchholz et al. 2018), the problem of limited sensitivity may remain across some populations (NICE 2025). As a solution, it was proposed to supplement CDs with additional dimension(s), referred to as *bolt-ons* (BOs), to offer more focus on specific HRQoL aspects (Mulhern et al. 2022). EuroQol is developing the EQ-5D Bolt-on Toolbox, an experimental set of bolt-ons to be used with the EQ-5D-5L in specific populations (EuroQol 2026; Rencz et al. 2026).

However, the introduction of BOs poses the following challenge, key to the motivation behind the present paper. To use information obtained with a BO in CUA, a value set must be available that relates the QALY weights to the BO levels. This triggers the question about the comparability of two sets of CUA results obtained with two different value sets: the core value set (using only CDs) and the BO value set (i.e., using jointly CDs and BO). The *raison d'être* of BOs is to measure HRQoL more precisely and, consequently, to affect CUA results. However, this raises a question of fairness. Technologies competing for the same resources may be evaluated using different value sets. A common worry is that using BOs may result in double-counting of health benefits as some HRQoL changes may be captured by both the CDs and the BOs, which can inflate the estimated QALY gains resulting from HRQoL improvement.

So far, the valuation-related research for BOs has focused on applying various preference elicitation methods, mostly time trade-off (TTO) and discrete choice experiment (DCE) (Geraerds et al. 2021; Rencz et al. 2025), not on studying consistency of value sets. Admittedly, when eliciting preferences for the extended instrument, i.e., for EQ-5D-5L with BO appended, CDs may receive less weight than in situations when BO is not included, which would reduce the double-counting presented above. The scaling factor approach (SFA) explicitly uses the empirical observation that CDs may receive lower weights when a BO is included (Yang et al. 2024). In models of TTO or DCE data for the extended instrument, the CD disutility coefficients are borrowed from the core value set. They are then uniformly scaled by an estimated parameter, λ, which is expected to be < 1.

However, in the absence of accepted definitions of double counting and consistency between value sets, it is not clear whether this preference-based scaling is of exactly the magnitude required to achieve consistency. Thus, although valuing the six-dimensional descriptive system may reduce the weights assigned to the CDs, this adjustment does not necessarily eliminate double counting or ensure consistency with the existing five-dimensional value set. Scaling in the SFA is based solely on preference-elicitation tasks, in which respondents consider hypothetical health states. By contrast, the extent of potential double counting in CUA also depends on the empirical association between CDs and BOs, that is, on the combinations of CD and BO levels observed in measurement studies.

Finally, it is being debated whether using λ in SFA is just an estimation step (i.e., whether to treat λ only as a nuisance parameter) or an inherent element of the final value set specification. Should λ be used only during estimation of the BO disutilities, it would be removed when producing the final BO value set, effectively setting $\lambda = 1$. Alternatively, the CD disutility coefficients could remain scaled by λ in the final value set. Our paper addresses these challenges as presented below.

We propose a novel and mathematically precise definition of consistency between the core and BO value sets that uses the measurement data for CDs and BO. We motivate why our definition captures what can be seen as fair or unfair differences between value sets. We then proceed to show how our approach can be used empirically, i.e., how to adjust an existing BO value set to make it consistent as per the proposed definition with an existing core value set. Finally, we illustrate the properties of the proposed method using both artificial and empirical data. The artificial data examples serve to illustrate more clearly the idea of the method in some admittedly unrealistic yet easy-to-read scenarios. The empirical data examples serve to verify whether the method can be used in real-life scenarios and whether the outcomes offer results that have face-validity.

It is important to stress that our paper is focused on establishing the consistency of value sets and not on the elicitation of preferences for BOs. We assume that disutility weights for CD and BO have been elicited and we demonstrate how to combine them to produce a BO value set that is consistent with a given core value set, while using our notion of consistency. In this sense, our paper is complementary to the previous research.

The problem of consistency is not limited to BOs. In CUA, different preference-based measures may be used,

including generic instruments such as the SF-6D and HUI and condition-specific instruments such as the EORTC QLU-C10D. The choice of instrument and associated value set may affect estimated QALY gains and incremental cost-effectiveness ratios (Davis et al. 2012; Grutters et al. 2007; Kim et al. 2021), thereby creating similar concerns about comparability and the potential for strategic instrument selection. An approach conceptually related to the one proposed here might be developed to assess the consistency of value sets from different instruments using measurement data in which the instruments are administered jointly. Such an extension is, however, beyond the scope of the present paper.

The paper is structured as follows. In Section "The proposed approach", we present the general idea of the approach, the notation to make the presentation rigorous, and the details of how the approach can be implemented. In Section "Simple illustrative examples", we present the properties of the method using simplified, clear-to-read examples. In Section "Empirical application", we apply the method to some actual BOs and datasets to further demonstrate the properties and the expected results of our approach. In Section "Discussion", we discuss our findings, focusing on the limitations of the approach, open questions, and possible ways forward. Finally, we provide some brief overall conclusions.

The proposed approach

Non-technical overview of the method

We compare the core and BO value sets using measurement data that contain both the CDs and the BO. If the BO value set assigns additional disutility mainly to health problems that tend to accompany problems in the CDs, we contract the BO value set by multiplying all its disutility coefficients by a common factor. This factor is chosen so that the two value sets satisfy the proposed consistency criterion, while preserving the ability of the BO to capture additional health problems.

Three cases help explain the problem. First, if BO levels are independent of the core profile, the BO provides information that cannot be predicted from the CDs, and there is no apparent reason to contract its contribution. Second, if the BO and the core profile convey exactly the same information, allowing the BO value set to produce larger QALY gains would amount to counting the same health differences twice. The most realistic case lies between these extremes: BO problems are associated with, but not fully determined by, problems in the CDs.

The remainder of this section first defines the core and BO value sets, then introduces the consistency criterion, and finally explains how the contraction factor is obtained.

Value sets

Value sets are functions that assign numbers to all health states in a domain defined by a given descriptive system. The values are based on the preference elicitation tasks conducted in a sample of respondents. For EQ-5D instruments, variants of TTO and DCE are typically used for preference elicitation, in isolation or together (Golicki et al. 2019; Ludwig et al. 2018; Norman et al. 2023; Rencz et al. 2020). For practical reasons, the preferences can be observed for only a subset of EQ-5D-5L states, and they are subsequently extrapolated to all 3125 states using econometric modelling. The usual specification of such models, which we assume in the paper, only considers additive impact of individual dimensions. Let $\delta(Q)$ denote the disutility of health state Q, defined as $\delta(Q) = 1 - u(Q)$. Under the additive specification assumed here, total disutility is the sum of the disutilities associated with the individual dimensions (see Roudijk et al., 2022, for a thorough review of EQ-5D-5L value-set methodologies Roudijk et al. 2022). More formally,

$$\delta(Q) = \alpha_{MO2} \times [MO(Q) = 2] + \cdots + \alpha_{MO5} \times [MO(Q) = 5] + \cdots + \alpha_{AD5} \times [AD(Q) = 5] \quad (1)$$

where

- αs are the parameters describing the importance of individual dimensions and levels (in the lower indices) and which are estimated based on observed preferences,
- $[\cdot]$ is a function that yields 1 if the bracketed condition is met, and 0 otherwise,
- MO(Q), ..., AD(Q) are functions that return the level of a given dimension for Q.

In the paper, we also assume the additive specification for BO value sets:

$$\delta^*(Q) = \alpha^*_{\mathrm{MO2}} \times [\mathrm{MO}(Q) = 2] + \ldots + \alpha^*_{\mathrm{BO2}} \times [\mathrm{BO}(Q) = 2] + \ldots \quad (2)$$

where the * superscript indicates that we refer to the BO value set, and BO(Q) denotes the BO dimension for state Q. Obviously, for the BO value set, Q is defined using six dimensions, i.e., EQ-5D-5L with BO appended, but to simplify we use the same symbol, with a slight abuse of notation (the context indicates whether Q is defined using 5 or 6 dimensions). Whenever referring to BO health states, we also use the + symbol to facilitate reading: for instance, the

worst BO health state (i.e., the new pits state) is denoted as $55555 + 5$.

Consistency criteria, an overview

The consistency of two value sets can be considered on various grounds. One could require *procedural consistency*, i.e., that two value sets should be produced using identical approaches: for instance, the same preference elicitation task, the same model specification, sampling from the same general population, or even them being produced within a single randomized study.

Alternatively, one could require *raw value-set consistency*, defined as some form of agreement between the disutility values or between the coefficients α and α^*. Several criteria are possible. The two value sets could span the same range, so that $\delta(55555) = \delta^*(55555 + 5)$.[1] Alternatively, appending a BO at level 1 could be required to have no effect: $\delta(Q) = \delta^*(Q + 1)$. A third possibility is to preserve the relative importance of the CDs: $\alpha_{\text{MO5}}/\alpha_{\text{SC5}} = \alpha^*_{\text{MO5}}/\alpha^*_{\text{SC5}}$, ...

In the present paper, we argue for a yet different approach to defining consistency, starting from the following motivation. The underlying reason to consider consistency of value sets is the fairness of decisions made with these value sets. Meanwhile, the value sets impact the decisions through the aggregated QALY value that results from applying the value sets to the distribution of health states in a given population. Hence, we argue the consistency criteria should be defined looking at the distribution of utility values produced by value sets in question for a given distribution of EQ-5D-5L + BO health states. The distribution of health states in a given population can be obtained from data on measurement of HRQoL using EQ-5D-5L with BO (for an overview of topics related to using EQ-5D instruments in measurement, see Devlin et al. 2020). Therefore, we refer to our approach as *value set consistency with measurement data*. We now proceed to some specific examples of such consistency criteria.

For instance, one could require that for a given distribution of Q, the average disutility (or utility, equivalently) should be the same with both value sets, i.e.,

$$\mathbb{E}(\delta(Q)) = \mathbb{E}(\delta^*(Q)) \tag{3}$$

(all expectations in the paper are taken with respect to the empirical distribution of Q). The above condition can be motivated by requiring either that (1) for both value sets prolonging the life of all individuals by the same amount should be assigned the same benefit or that (2) for both value sets the benefit of treating all individuals to perfect health should be identical.

Alternatively, one could require that the standard deviation (SD) of disutility should be the same for both value sets, i.e.,

$$\mathbb{SD}(\delta(Q)) = \mathbb{SD}(\delta^*(Q)). \tag{4}$$

The above condition can be motivated as follows. SD measures the dispersion of values, hence, it is also an approximate measure of differences between individual values (for normal distribution, explicit formula links SD and mean absolute deviation). Assuming that the sample consists of individuals in various health state severity, SD informs about the size of possible improvements from changing the health state from one state to another.

However, the above notions of consistency do not account for the rationale of including BO, i.e., the above criteria strive to completely eradicate the impact of a BO and so they eradicate any effect of increased measurement sensitivity. Meanwhile, retaining some impact of BO may be warranted. For instance, assume an extreme case of a complete ceiling effect, i.e., all CDs being at level 1, such that all individuals are assigned disutility equal to 0 with the core value set. If BO captures HRQoL problems in this population, this would provide just the right rationale for using the BO. However, since it would also result in strictly positive disutility values with the BO value set, the above consistency criteria cannot be achieved, unless one altogether disregards BO in the calculation of δ^*. In the following subsection, we propose a potential fair middle ground.

The proposed consistency criterion

In measurement data, health problems reported in a BO are often associated with problems reported in the CDs (see the bottom parts of Tables 5 and 6). The proposed criterion uses the conditional distribution of BO levels given the core profile to represent this association.

Consider a technology that improves the BO only to the extent that would be expected from the accompanying improvements in the CDs under this observed conditional relationship. In such a case, the BO does not reveal an additional treatment effect beyond that already associated with the core profile. It therefore seems undesirable for the estimated QALY gain to depend on whether the core or the BO value set is used. A larger gain obtained solely by selecting the BO value set could reflect strategic value-set selection rather than greater measurement sensitivity.

[1] It is more convenient to work with disutilities than with utilities, and there is a trivial correspondence between the two. In this paper, we leave aside the terminological discussion whether the numbers collected in value sets should be referred to as utilities, in view of how they are derived from the aggregation of individuals' preferences (Jakubczyk 2023).

We therefore require the two value sets to assign the same magnitude of benefit to changes in the core profile accompanied by BO changes that follow the conditional distribution observed in measurement data. We operationalize this magnitude as the standard deviation of expected disutility across core health profiles, as formalized below.

To formalize this idea, we compare the distributions of $\mathbb{E}(\delta(Q) \mid \text{core})$ and $\mathbb{E}(\delta^*(Q) \mid \text{core})$, where $\text{core} = \{\text{MO}, \ldots, \text{AD}\}$. The first distribution describes how disutility under the core value set varies across core health profiles. The second describes the corresponding expected disutility under the BO value set, averaging over the distribution of BO levels observed for each core profile. For consistency, we require these two distributions to be similar. To operationalize this similarity, we use their standard deviations, as motivated above.

Eventually, we propose to use the following consistency criterion:

$$\mathbb{SD}(\mathbb{E}(\delta(Q)|\text{core})) = \mathbb{SD}(\mathbb{E}(\delta^*(Q)|\text{core})), \tag{5}$$

i.e. the equality of SD of conditional expected disutility with conditioning being with respect to the CDs. In intuitive terms, the criterion requires the two value sets to produce the same degree of variation in expected disutility across core health profiles. Thus, moving between core profiles, together with the BO differences typically associated with those profiles, should generate QALY differences of the same overall magnitude under either value set. In Section "Simple illustrative examples", we illustrate the intuitive properties of this approach in a simplified, easy-to-read setting.

The proposed criterion is asymmetric by design. We treat the existing core EQ-5D-5L descriptive system and its value set as the reference standard and ask how an extended EQ-5D-5L+bolt-on value set can be made consistent with that standard. This does not imply that the CDs are mutually independent or that overlap between them cannot exist. Rather, any such overlap is already embedded in the established descriptive system and value set. Our purpose is not to re-optimise or re-scale the core EQ-5D-5L, but to prevent the addition of a BO from introducing further changes that are attributable merely to information already represented in the CDs.

Having defined consistency, the next step is to determine how an existing BO value set can be adjusted to satisfy this criterion.

Obtaining consistency by contraction

Both the expectation and the standard deviation are homogeneous with respect to multiplication by a positive scalar. Therefore, multiplying $\delta^*(Q)$ by a constant κ multiplies the standard deviation of its conditional expectation by the same constant:

$$\mathbb{SD}\left(\mathbb{E}(\kappa \times \delta^*(Q) \mid \text{core})\right) = \kappa \times \mathbb{SD}\left(\mathbb{E}(\delta^*(Q) \mid \text{core})\right).$$

This result provides a direct way to obtain consistency. We choose κ so that Eq. 5 is satisfied. The interpretation is straightforward: if the unadjusted BO value set produces greater variation across core profiles than the core value set, $\kappa < 1$ reduces this variation proportionally. A value of $\kappa \approx 1$ indicates that little adjustment is needed, whereas a smaller value indicates that a larger adjustment is required because of the combination of information overlap and the magnitude of BO disutility.

To avoid terminological confusion with SFA, we refer to multiplying δ^* by κ as *contraction*, as we expect $\kappa < 1$ in real-life situations (confirmed by our empirical application reported below).

To examine the contraction mechanism separately from differences in the CD coefficients, we next consider a simplified construction of δ^*. We construct δ^* by appending the BO terms to δ. We retain the same coefficients for all CDs. Thus, $\alpha_{\text{MO2}} = \alpha^*_{\text{MO2}}$, ..., $\alpha_{\text{AD5}} = \alpha^*_{\text{AD5}}$. Such δ^* could be obtained, for instance, by setting $\lambda = 1$ in SFA in the construction of the final BO value set. In this way, we can examine the contraction required to guarantee consistency as defined in Eq. 5 as a function of the association between the core profile and the BO, without introducing differences between the CD coefficients in $\delta(\cdot)$ and $\delta^*(\cdot)$. Neither the consistency criterion nor the possibility of satisfying it by contraction depends on this simplified form of $\delta^*(\cdot)$.

In the main approach presented in the paper, we assume that the entire function $\delta^*(\cdot)$ is contracted including the BO part, i.e., that all the coefficients, α^*_{MO2}, ..., α^*_{BO5}, are contracted using the same κ. However, as a form of sensitivity analysis, we also consider other approaches in the empirical illustration presented in Section "Empirical application".

Simple illustrative examples

In this section, we use simplified examples to illustrate the intuitive properties of the proposed contraction method. To make the calculations easy to follow, we assume that the descriptive system contains only one CD and one BO. The examples differ in the strength of the association between the CD and the BO in measurement data.

We begin with two limiting cases. In Example A, the BO is independent of the CD and therefore provides information that cannot be predicted from the CD. In Example B, the BO and the CD are perfectly positively associated and

provide no distinct information in the observed population. We then consider Example C, the more realistic intermediate case of positive but imperfect association. This ordering shows how the method moves from no contraction to stronger contraction as the information represented by the BO increasingly overlaps with that represented by the CD.

In all three examples, we use the same disutility coefficients for the CD:

$$\alpha_{CD2} = 0.1, \quad \alpha_{CD3} = 0.2, \quad \alpha_{CD4} = 0.4, \quad \alpha_{CD5} = 0.6,$$

and for the BO:

$$\alpha_{BO2} = 0.05, \quad \alpha_{BO3} = 0.1, \quad \alpha_{BO4} = 0.2, \quad \alpha_{BO5} = 0.3.$$

We also use the same marginal distributions of CD and BO levels in all examples. For instance, 30% of the population is at level 1 of the CD.

Example A: independence

Example A illustrates the limiting case in which BO levels are independent of the CD level. Its purpose is to demonstrate that the method leaves the BO value set unchanged when the BO provides information that cannot be predicted from the CD.

The example is presented in Table 1. The upper part of the table shows the original, uncontracted values. The first three columns report the CD level, its population frequency, and the corresponding disutility under the core value set. The next five columns show the joint population distribution across BO levels within each CD level. The following five columns report disutilities under the uncontracted BO value set, δ^*. The final column presents the mean of δ^* conditional on the CD level.

Because the BO is independent of the CD, the conditional distribution of BO levels is the same for every CD level. Consequently, the expected BO contribution is constant across CD levels. Adding the BO therefore shifts conditional expected disutility by the same amount for every CD level (by 0.085) but does not change its dispersion. The proposed consistency criterion is therefore already satisfied by the uncontracted BO value set, and the required contraction parameter is trivially $\kappa = 1$.

The lower part of Table 1 presents the values after applying the contraction parameter. Because $\kappa = 1$, all disutilities remain unchanged. This is not merely a numerical coincidence: it reflects the fact that the BO contribution is unrelated to the CD level and therefore represents additional information rather than potential double counting.

The key takeaway from Example A is that no contraction is required when the BO is independent of the CD. The

Table 1 Illustration of the contraction method for a case of independence between a single core dimension (CD) and a bolt-on (BO) in a hypothetical population (pop.)

Example A — independence, before contraction													
CD	% pop.	CD δ	% pop. for BO levels					δ^* for BO levels					mean δ^* by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	9	9	6	3	3	.0	.05	.1	.2	.3	.085
2	30	.1	9	9	6	3	3	.1	.15	.2	.3	.4	.185
3	20	.2	6	6	4	2	2	.2	.25	.3	.4	.5	.285
4	10	.4	3	3	2	1	1	.4	.45	.5	.6	.7	.485
5	10	.6	3	3	2	1	1	.6	.65	.7	.8	.9	.685
Example A cont'd — with trivial contraction using $\kappa = 1$													
CD	% pop.	CD δ	% pop. for BO levels					$\kappa\delta^*$ for BO levels					mean $\kappa\delta^*$ by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	9	9	6	3	3	.0	.05	.1	.2	.3	.085
2	30	.1	9	9	6	3	3	.1	.15	.2	.3	.4	.185
3	20	.2	6	6	4	2	2	.2	.25	.3	.4	.5	.285
4	10	.4	3	3	2	1	1	.4	.45	.5	.6	.7	.485
5	10	.6	3	3	2	1	1	.6	.65	.7	.8	.9	.685

Note: Population percentages in the five BO-level columns are joint percentages of the total population. The final column gives the mean BO value-set disutility conditional on the CD level. Calculations use the unrounded value of κ and unrounded intermediate quantities; displayed values are rounded

method retains the BO disutilities in full because the BO adds variation that is not associated with differences in the core profile.

Example B: perfect positive association

Example B illustrates the opposite limiting case, in which the BO and the CD are perfectly positively associated. Its purpose is to show what the method does when the BO does not distinguish between individuals beyond what is already captured by the CD.

The example is presented in Table 2. The table has the same structure as Table 1: the upper part reports the uncontracted values, whereas the lower part shows the values after applying the contraction parameter.

Because each CD level is always accompanied by the corresponding BO level, the BO adds no independent information in the observed population. The uncontracted BO value set nevertheless assigns greater disutility differences between CD levels than the core value set, because it adds BO disutility to health problems that are already fully represented by the CD (or at least can be fully predicted by CD). Hence, contraction is needed, and the required contraction parameter is $\kappa \approx 0.667$.

After contraction, the effect of adding the BO coefficients is fully neutralized. For every health state observed in the population, the disutility under the contracted BO value set is identical to that under the core value set.

The key takeaway from Example B is that, when the CD and the BO are perfectly associated and their coefficients are proportional, contraction completely removes the additional effect of the BO. The contracted BO value set then reproduces the core value-set disutilities for the observed health states.

Taken together, Examples A and B establish the two intuitive boundary cases. Independence leads to no contraction, whereas complete overlap leads to full neutralization of the additional BO contribution in the observed population.

Example C: positive but imperfect association

Example C illustrates how the contraction method operates in the more realistic case in which the BO and the CD are positively, but not perfectly, associated. Its purpose is to show how contraction reduces the additional variation generated by the BO value set without eliminating the BO's ability to distinguish between health states.

The example is presented in Table 3. The table has the same structure as Tables 1 and 2: the upper part shows the original, uncontracted values, whereas the lower part presents the values after contraction.

Table 2 Illustration of the contraction method for a perfect association between a single core dimension (CD) and a bolt-on (BO) in a hypothetical population (pop.)

Example B — perfect positive association, before contraction												
CD	% pop.	CD δ	% pop. for BO levels					δ^* for BO levels					mean δ^* by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	30	0	0	0	0	.0	.05	.1	.2	.3	.00
2	30	.1	0	30	0	0	0	.1	.15	.2	.3	.4	.15
3	20	.2	0	0	20	0	0	.2	.25	.3	.4	.5	.30
4	10	.4	0	0	0	10	0	.4	.45	.5	.6	.7	.60
5	10	.6	0	0	0	0	10	.6	.65	.7	.8	.9	.90
Example B cont'd — with contraction using $\kappa \approx .667$													
CD	% pop.	CD δ	% pop. for BO levels					$\kappa\delta^*$ for BO levels					mean $\kappa\delta^*$ by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	30	0	0	0	0	.00	.03	.07	.13	.20	.0
2	30	.1	0	30	0	0	0	.07	.10	.13	.20	.27	.1
3	20	.2	0	0	20	0	0	.13	.17	.20	.27	.33	.2
4	10	.4	0	0	0	10	0	.27	.30	.33	.40	.47	.4
5	10	.6	0	0	0	0	10	.40	.43	.47	.53	.60	.6

Note: Population percentages in the five BO-level columns are joint percentages of the total population. The final column gives the mean BO value-set disutility conditional on the CD level. Calculations use the unrounded value of κ and unrounded intermediate quantities; displayed values are rounded

There is a positive association between the CD and the BO, with a Spearman correlation coefficient of approximately 0.84. Using the uncontracted BO value set instead of the core value set produces a larger range of disutilities: $\delta^*(5+5) = 0.9 > 0.6 = \delta(5)$. It also produces a larger average disutility, a larger SD of disutility, and a larger SD of conditional expected disutility.

The required contraction parameter is

$$\kappa = \frac{\mathbb{SD}\big(\mathbb{E}(\delta(Q) \mid \mathrm{CD})\big)}{\mathbb{SD}\big(\mathbb{E}(\delta^*(Q) \mid \mathrm{CD})\big)} \approx 0.699.$$

The full calculation is presented in the Supplementary Materials.

After contraction, the standard deviation of conditional expected disutility is the same under the core value set and the contracted BO value set. The BO nevertheless continues to assign a positive QALY gain to improvements in the BO. Thus, the method does not remove the additional sensitivity provided by the BO.

At the same time, an improvement in the CD alone receives a smaller QALY gain under the contracted BO value set than under the core value set. This reduction reflects the fact that changes in the CD are typically accompanied by some changes in the BO. Contracting both sets of coefficients reduces the risk that these associated changes are counted twice.

The key takeaway from Example C is that, under positive but imperfect association, contraction produces an intermediate result. It reduces the impact of overlapping information while preserving the additional sensitivity provided by the BO. Thus, Example C lies between the two boundary cases: unlike Example A, some contraction is required, but unlike Example B, the additional contribution of the BO is not fully neutralized.

Table 3 Illustration of the contraction method for a positive but imperfect association between a single core dimension (CD) and a bolt-on (BO) in a hypothetical population (pop.)

Example C — positive but imperfect association, before contraction													
CD	% pop.	CD δ	% pop. for BO levels					δ^* for BO levels					mean δ^* by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	20	10	0	0	0	.0	.05	.1	.2	.3	.017
2	30	.1	10	10	10	0	0	.1	.15	.2	.3	.4	.150
3	20	.2	0	10	5	5	0	.2	.25	.3	.4	.5	.300
4	10	.4	0	0	5	3	2	.4	.45	.5	.6	.7	.570
5	10	.6	0	0	0	2	8	.6	.65	.7	.8	.9	.880
Example C cont'd — with contraction using $\kappa \approx .699$													
CD	% pop.	CD δ	% pop. for BO levels					$\kappa\delta^*$ for BO levels					mean $\kappa\delta^*$ by CD level
			1	2	3	4	5	1	2	3	4	5	
1	30	.0	20	10	0	0	0	.00	.03	.07	.14	.21	.012
2	30	.1	10	10	10	0	0	.07	.10	.14	.21	.28	.105
3	20	.2	0	10	5	5	0	.14	.17	.21	.28	.35	.210
4	10	.4	0	0	5	3	2	.28	.31	.35	.42	.49	.399
5	10	.6	0	0	0	2	8	.42	.45	.49	.56	.63	.615

Note: Population percentages in the five BO-level columns are joint percentages of the total population. The final column gives the mean BO value-set disutility conditional on the CD level. Calculations use the unrounded value of κ and unrounded intermediate quantities; displayed values are rounded

Empirical application

Measurement data used

We test our method separately for 9 BOs: breathing (BR), cognition (CG), self-confidence (CO), hearing (HE), skin irritation (SI), sleep (SL), social relationships (SR), tiredness (TI), and vision (VI). These BOs were selected because they are the most commonly used BOs, and 8 of them were also selected to be included in the experimental EQ-5D Bolt-on Toolbox (however, the BO wording used in the included studies slightly differed from that in the experimental EQ-5D Bolt-on Toolbox for all dimensions except SI and CO).

The required contraction parameter κ is obviously driven by the empirical distribution of health states in the considered population. In this paper, we study two different types of population: the general population (which is irrespective of the particular BO in question) and the patient population (which depends on what BO is being considered). The wording of BOs in all datasets selected for this study was identical, and available from the EuroQol Office.

The general-population dataset pooled data from three sources: the EQ-DAPHNIE survey project (Johnson et al. 2025; Dewilde et al. 2025), and Rencz and Janssen 2024. It included 46,774 respondents from 14 countries with complete information on the CDs. Depending on the BO, the available sample size ranged from 10,395 to 42,206. Further details, including BO-specific sample sizes and country composition, are provided in the Supplementary Materials.

The patient dataset pooled observations from studies involving breast cancer (Pangestu et al. 2025), coeliac disease (Angyal et al. 2025), chronic obstructive pulmonary disease (Adab et al. 2017; Hoogendoorn et al. 2021), myasthenia gravis (Dewilde et al. 2024), psoriasis (Swinburn et al. 2013), and minimally invasive cosmetic procedures (Müller et al. 2025). It included 3,562 respondents with complete information on the CDs. For each BO-specific analysis, we included only respondents with complete information on all five CDs and the relevant BO. Depending on the BO, the available sample size ranged from 312 to 2,486; no patient data were available for hearing. Further dataset details are provided in the Supplementary Materials.

Disutility values used

To define the core value set, $\delta(\cdot)$, we use disutility coefficients coming from the recent UK EQ-5D-5L valuation study (Rowen et al. 2026). The disutility values for BOs were based on the available literature, with some modifications as deemed necessary and as described below. Ideally, the CD and BO disutility coefficients would be obtained from valuation studies conducted in the same country and target population. This was not possible in the present application because country-specific BO coefficients were available for only a limited number of dimensions. We therefore used a single core value set for all BOs and combined it, where necessary, with BO coefficients obtained in other countries. This choice facilitates comparison across BOs by holding the core value set constant, but it may affect the estimated contraction parameters if the relative importance assigned to the CDs and the BO differs across countries. Accordingly, the resulting values of κ should be interpreted as methodological illustrations rather than country-specific recommendations for final BO value sets.

For BR, we used data from Table 4 in a study by Hoogendoorn et al. (2019). We replaced the negative disutility coefficient for level 2 with 0. For CG, we used data from a study by Yang et al. (2026), Table 3 (Model_CO). We transformed the incremental dummies into cumulative dummies while replacing negative incremental dummies with 0s. Coefficients for CO and SI were based on the results of Swinburn et al. (2013), Table 4 (Model 4), where reported coefficients for some neighbouring levels were set as identical to avoid non-monotonicity.

For HE and TI, we used results of Yang et al. (2015) in Table 4 (Model 1). That study did not differentiate the coefficients associated with BO levels between HE and TI (the difference between BOs was captured by another variable, which was not level-specific). Moreover, the study was conducted for EQ-5D-3L, so we assumed that level 2 (level 3) of EQ-5D-3L corresponds to level 3 (level 5) of EQ-5D-5L, and we used linear interpolation to fill in the values for levels 2 and 4 of EQ-5D-5L. For SL, we used a study by Kharroubi et al. (2020), Table 2 (RE model, B part of the table), with an analogous approach to convert values from three to five levels and replacing negative disutilities with 0.

For SR, we used unpublished results of Yang et al. (2024). For VI, we used results of Yang et al. (2023), Table 2 (EQ + VI), while averaging coefficients for levels 2 and 3 to avoid inconsistency.

The disutility values for all dimensions and levels are presented in Table 4.

Alternative approaches to calculating κs

We next examine whether the results are robust to alternative definitions of consistency and alternative ways of applying contraction. In the main approach, we use Eq. 5 to define consistency and contract the whole function $\delta^*(\cdot)$ in Eq. 2. As a form of sensitivity analysis, we report the following four other approaches. First, we contract only the part of $\delta^*(\cdot)$ corresponding to CD.

Additionally, instead of using the SD of conditional expectation (i.e., Eq. 5) to define consistency, we use either expected value (Eq. 3) or SD (Eq. 4). In both of these approaches, we contract the whole function $\delta^*(\cdot)$.

Finally, we attempt to acknowledge varying correlation between BO and individual CDs by varying the contraction of each of the CDs in δ^*. To do this, we apply the following approach. First, based on Eq. 5, we calculate CD-specific κ assuming that only coefficients related to this CD are contracted in Eq. 2. Denote these contraction coefficients by κ_{MO}, ..., κ_{AD}. We look for a value of the parameter ρ such that scaling individual CDs with, respectively, $\kappa_{\text{MO}}^{\rho}$, ..., $\kappa_{\text{AD}}^{\rho}$, leads to the consistency as per Eq. 5 (values $\rho < 1$

Table 4 Disutility values for core and bolt-on dimensions (non-incremental, i.e., should not be acummulated). MO = mobility, SC = self-care, UA = usual activities, PD = pain & discomfort, AD = anxiety & depression, VI = vision, HE = hearing, BR = breathing, SL = sleep, TI = tiredness, SR = social relationships, SI = skin irritation, CO = self-confidence, CG = cognition

	Core dimensions					Bolt-ons							
Level	MO	SC	UA	PD	AD	BR	CG	CO	HE	SI	SL	SR	TI	VI
2	.032	.038	.049	.056	.041	0	.048	.016	.0075	.052	0	.038	.0075	.022
3	.058	.060	.086	.066	.126	.086	.048	.016	.015	.052	0	.038	.015	.022
4	.179	.162	.184	.371	.313	.219	.206	.084	.064	.103	.0185	.160	.064	.159
5	.279	.206	.212	.479	.391	.327	.385	.084	.113	.123	.037	.219	.113	.411

are expected, as less contraction is needed when all CDs are contracted, rather than only one CD).

Results for the general population

The results for the general population are presented in Table 5, both for the main approach and for the additional approaches specified in Section "Alternative approaches to calculating κs". The required contraction for the main approach ranges 0.905–0.986, and typically a stronger contraction is needed in alternative approaches. Obviously, the contraction reduces the disutility of the pits state (55555+5) as compared with the uncontracted disutility δ^*. For all BOs, the contracted disutility of 55555+5 is larger than the disutility of the CD-only pits (55555), so the range of possible (dis)utility values is larger for the contracted BO value set than for the core value set. When CD-varying contractions are considered in one of the additional approaches, typically a stronger contraction is required for AD, which corresponds to this CD being the one most strongly correlated with the majority of BOs.

In the upper part of Fig. 1, we illustrate how the contraction parameter, κ, changes with the disutility of BO (for level 5) and the strength of correlation with CDs (the maximal Spearman correlation coefficient across CD). In the regression analysis, both variables are negatively associated with κ: for the level-5 disutility, the slope = −0.19 (p-value = 0.006), while for the correlation the slope = −0.13 (however, non-significant, with p-value = 0.077). The presented regression line is a slice of a multivariate regression surface for the average value of maximal Spearman correlation (0.41). Additionally, in the Supplemental Materials in Fig. S1, we present the results for the independent variables swapped (i.e., correlation presented on the X axis, and disutility presented using shade of gray).

The stronger observed association of κ with BO disutility than with the correlation measure does not imply that correlation is unimportant to the proposed consistency criterion. The two variables play complementary roles. Correlation reflects the extent to which the information provided by the BO overlaps with information already captured by the CDs, whereas BO disutility determines the numerical importance of that overlapping information in the value set. Consequently, little or no contraction is required if either component is absent: a BO with zero disutility cannot affect calculated utilities, irrespective of its correlation with the CDs, while a BO that is independent of the CDs adds new information and therefore does not create double counting, irrespective of its disutility. The required contraction is thus jointly determined by the strength of association in measurement and by the preference weight assigned to the BO.

Table 5 General population (MO = mobility, SC = self-care, UA = usual activities, PD = pain/discomfort, AD = anxiety/depression, BR = breathing, CG = cognition, CO = self-confidence, HE = hearing, SI = skin irritation, SL = sleep, SR = social relationships, TI = tiredness, VI = vision). κ denotes the contraction coefficient; δ denotes the disutility for CD; δ^* denotes the disutility of BO value set prior to contraction

	BR	CG	CO	HE	SI	SL	SR	TI	VI
Main approach, equating conditional standard deviation wrt. core dimensions									
κ	0.922	0.905	0.954	0.980	0.950	0.986	0.905	0.945	0.936
$\delta(55555)$	1.567	1.567	1.567	1.567	1.567	1.567	1.567	1.567	1.567
$\delta^*(55555+5)$	1.894	1.952	1.651	1.680	1.690	1.604	1.786	1.680	1.978
$\kappa \times \delta^*(55555+5)$	1.746	1.767	1.575	1.646	1.606	1.582	1.616	1.588	1.851
Alternative approaches to contraction									
core dimensions only, based on conditional SD	0.915	0.895	0.952	0.980	0.948	0.986	0.894	0.942	0.931
all dimensions, based on expected disutility	0.941	0.880	0.915	0.971	0.874	0.989	0.851	0.921	0.864
all dimensions, based on unconditional SD	0.914	0.895	0.951	0.979	0.943	0.986	0.895	0.942	0.925
Varying contraction per core dimension, based on conditional SD wrt. core dimensions									
MO	0.918	0.929	0.976	0.981	0.956	0.989	0.940	0.956	0.947
SC	0.895	0.886	0.964	0.972	0.941	0.987	0.914	0.953	0.931
UA	0.914	0.900	0.960	0.980	0.951	0.986	0.911	0.944	0.941
PD	0.923	0.911	0.969	0.980	0.950	0.987	0.926	0.944	0.937
AD	0.922	0.876	0.922	0.981	0.943	0.983	0.833	0.929	0.927
Additional information, Spearman correlation with core dimensions									
MO	0.33	0.27	0.14	0.29	0.19	0.28	0.18	0.29	0.22
SC	0.32	0.28	0.19	0.31	0.22	0.26	0.24	0.25	0.18
UA	0.37	0.39	0.26	0.28	0.23	0.37	0.30	0.39	0.23
PD	0.34	0.32	0.26	0.30	0.29	0.42	0.27	0.44	0.30
AD	0.30	0.38	0.55	0.19	0.29	0.47	0.55	0.47	0.26

Overall, the general-population results indicate that only modest contraction is required for the BOs considered. The contracted BO value sets continue to span a wider range of disutility than the core value set, while satisfying the proposed consistency criterion. The amount of contraction reflects both the strength of the association between the BO and the CDs and the magnitude of the BO disutility coefficients. It also depends on the underlying core value set. To illustrate this, we repeated the calculations using the Polish EQ-5D-5L value set instead of the UK value set. The resulting contraction parameters were highly correlated with the original results (Pearson correlations of 0.990 and 0.994 for the general and patient populations, respectively), although they were consistently somewhat lower. Detailed results are presented in the Supplementary Materials.

Results for the patient population

The results for the patient population are presented in Table 6. Similar patterns prevail as for the general population. In the bottom part of Fig. 1, we illustrate how κ changes with the disutility of BO and the maximal Spearman correlation of BO with CDs (the slopes are, respectively, −0.20 and −0.13, with respective p-values of 0.025 and 0.246). In the figure, the presented regression line is a slice of a multivariate regression surface for the maximal Spearman correlation value of 0.45. In Fig. S1, we present the results with the independent variables swapped in terms of presentation.

Overall, the patient-population analysis produced contraction parameters broadly similar to those observed in the general population. This suggests that, for the BOs and datasets considered, the choice between general-population and patient measurement data has only a limited effect on the required contraction. This finding should nevertheless be interpreted cautiously because the patient samples differed across BOs and were small for some dimensions.

Discussion

We first summarize the principal findings and then discuss their implications, relationship to the SFA, limitations, and possible extensions.

General findings

In this paper, we proposed a new approach to provide consistency between two value sets: an existing core EQ-5D-5L value set and an EQ-5D-5L+BO value set being built. We suggest that the approach should be a part of the BO value set development process, but it can also be applied post-hoc for existing BO value sets. Our approach consists of a formal definition of consistency and a contraction of disutility coefficients that produces such a consistency. Put simply, the method asks how much of the additional variation introduced by a BO is already predictable from the core profile

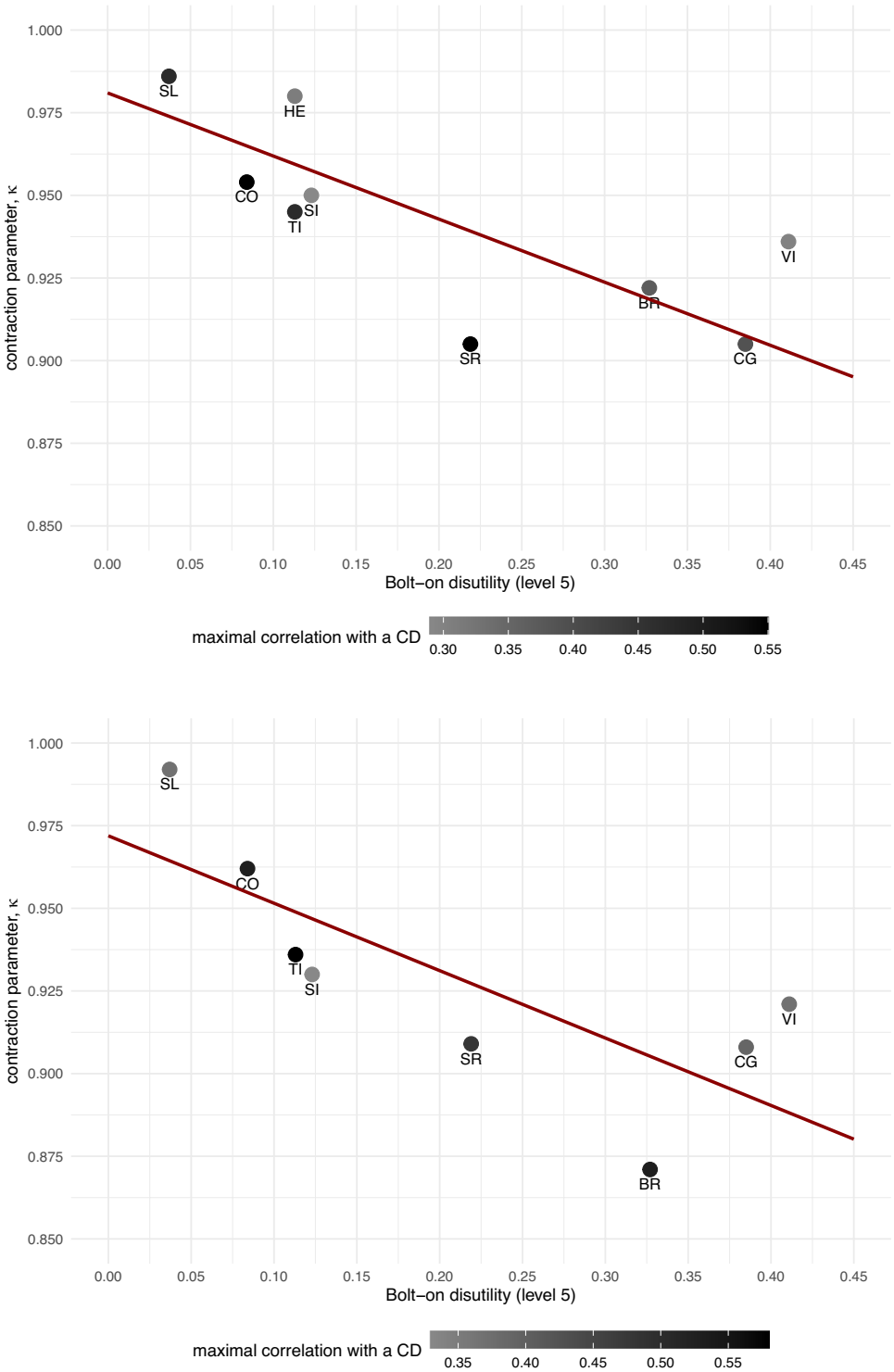


Fig. 1 Contraction parameter κ versus the bolt-on disutility (for level 5) and correlation with core dimensions (CD, maximal correlation across all CDs), results for the general population (upper part) and the patient population (bottom part). BR = breathing, CG = cognition, CO = self-confidence, HE = hearing, SI = skin irritation, SL = sleep, SR = social relationships, TI = tiredness, VI = vision. Regression line drawn for mean value of correlation

and then reduces the scale of the BO value set accordingly. The postulated definition takes into account the association between the BO and the CDs observed in the measurement of HRQoL. We argue in the paper that in this way we are able to eradicate an unfair impact of appending BO to CDs in the descriptive system, i.e., such an impact that effectively repeats information already contained in the CDs. Other things being equal, stronger association between the BO and the core profile generally requires stronger contraction to achieve consistency.

We believe that the use of our method may increase the acceptability of BOs, especially among the recipients of CUA results and in particular among those making reimbursement decisions based on such results. If core and BO

Table 6 Patient population (MO = mobility, SC = self-care, UA = usual activities, PD = pain/discomfort, AD = anxiety/depression, BR = breathing, CG = cognition, CO = self-confidence, SI = skin irritation, SL = sleep, SR = social relationships, TI = tiredness, VI = vision). κ denotes the contraction coefficient; δ denotes the disutility for CD; δ^* denotes the disutility of BO value set prior to contraction

	BR	CG	CO	SI	SL	SR	TI	VI
Main approach, equating conditional standard deviation wrt. core dimensions								
κ	0.871	0.908	0.962	0.930	0.992	0.909	0.936	0.921
$\delta(55555)$	1.567	1.567	1.567	1.567	1.567	1.567	1.567	1.567
$\delta^*(55555+5)$	1.894	1.952	1.651	1.690	1.604	1.786	1.680	1.978
$\kappa \times \delta^*(55555+5)$	1.650	1.772	1.588	1.572	1.591	1.623	1.572	1.822
Alternative approaches to contraction								
core dimensions only, based on conditional SD	0.850	0.898	0.960	0.925	0.992	0.900	0.932	0.913
all dimensions, based on expected disutility	0.875	0.858	0.927	0.828	0.991	0.895	0.912	0.886
all dimensions, based on unconditional SD	0.861	0.902	0.960	0.922	0.992	0.904	0.934	0.908
Varying contraction per core dimension, based on conditional SD wrt. core dimensions								
MO	0.818	0.985	0.976	0.908	0.994	0.913	0.931	0.931
SC	0.842	0.969	0.969	0.913	0.993	0.906	0.934	0.930
UA	0.830	0.863	0.968	0.926	0.992	0.900	0.923	0.898
PD	0.868	0.920	0.971	0.924	0.991	0.926	0.938	0.938
AD	0.899	0.875	0.928	0.939	0.989	0.870	0.932	0.933
Additional information, Spearman correlation with core dimensions								
MO	0.53	0.16	0.19	0.30	0.30	0.33	0.51	0.26
SC	0.38	-0.02	0.24	0.31	0.25	0.35	0.42	0.24
UA	0.46	0.39	0.25	0.32	0.35	0.44	0.58	0.37
PD	0.40	0.37	0.26	0.33	0.37	0.33	0.53	0.31
AD	0.21	0.37	0.53	0.23	0.35	0.47	0.42	0.21

value sets are shown to be consistent in an accepted sense, it may still be considered fair that CUA using BO value set produces more favourable results for a given technology.

What adds to the motivation for the present study is the large variability of disutility values for individual BOs reported in the literature (see Table 4): for some BOs, barely any disutility is associated (for instance, SL), while for some BOs, the disutility values exceed many of those reported for the CDs (for instance, BR and CG). Such variability is to be expected in future BO valuation studies in view of the large number of BOs, the large number of countries in which such studies could be conducted, and the likely smaller sample sizes compared with core instrument valuation studies. In view of this heterogeneity, it is even more important to regularize the ultimate BO value set in some way to make it consistent with core value sets.

The required contraction, i.e., κ parameter, is in the range of 0.905–0.986 for the general population data, and 0.871–0.992 for the patient population. The average difference in κ between the general population and patient population was only .009 (and .0145 for average absolute difference). For 4 BOs (BR, SI, TI, and VI), a stronger contraction was needed in patient population, and also for 4 BOs (CG, CO, SL, and SR), a stronger contraction was needed in general population. Somewhat surprisingly, the choice of population (general vs. patient) from which measurement data were drawn had little impact on our results. One possible explanation is that patient populations differ from the general population mainly in the marginal distribution of severity, with a larger proportion of observations at more severe levels, while the conditional distribution of BO responses given the core profile may remain broadly similar. A change in the marginal distribution can increase rank correlations by reducing the concentration of observations at the lowest levels, but it does not necessarily change κ to the same extent. The contraction parameter κ is based on the relative dispersion of core disutility and conditional expected BO disutility, rather than directly on a correlation coefficient. Examples D1 and D2 in the Supplementary Materials illustrate how two populations can have markedly different Spearman correlations but require virtually identical contraction when their underlying conditional relationship is preserved. This explanation does not imply that the choice between general-population and patient data is immaterial in general. Differences in the conditional distribution of BO levels given the core profile could lead to materially different contraction parameters in other populations. When more measurement data are available, this finding and explanation could be subject to further studies.

In the empirical part of our paper, a stronger contraction was needed for BOs with larger disutility levels (see Fig. 1), as such BOs affect the value sets more when appended to the descriptive system. This is intuitive and may be seen as a safety mechanism: if a BO valuation study overestimated the disutility values (say, simply due to the randomness of the sample used), a stronger contraction will somewhat reduce this overestimation.

Consistently with the conceptual basis of the method, BOs that were more strongly associated with the CDs required stronger contraction on average, although this association was not statistically significant in the regression analysis. This lack of significance may have resulted from low proportion of observations at levels 4 and 5 (see Tables S2 and S5 in the Supplementary Materials), unlike in our artificial examples in Section "Simple illustrative examples". Meanwhile, only at such levels are the disutility values substantially large.

Intuitively, when only CDs are subject to contraction (as in one of the approaches in the sensitivity analysis), a slightly stronger contraction is needed. A much stronger contraction is required, when the equality of average disutility (or equivalently, of utility) is used as a consistency criterion. Hence, when the main approach proposed in the paper is used, one should expect that the average disutility (respectively, utility) with the contracted BO value set will be greater (respectively, smaller) than the average disutility (respectively, utility) with the core value set. Admittedly, this may result in lower QALY gains from the life extension technologies obtained with the contracted BO value set. This may be treated as another safety mechanism to prevent the adoption of BO value sets for unjustified reasons.

When the contraction parameter was allowed to differ across CDs, the strongest contraction was typically needed for CDs that most strongly correlated with a BO in question. For instance, for CO in the patient population, AD was contracted more strongly ($\kappa_{AD} = 0.928$) than PD ($\kappa_{PD} = 0.971$). However, in some cases the contraction differed between CDs that correlated with the BO to a similar extent. For instance, for CG in the patient population, AD was contracted more strongly ($\kappa_{AD} = 0.875$) than PD ($\kappa_{PD} = 0.920$), in spite of the same value of Spearman correlation. We think that this reflects the limitations of Spearman correlation in measuring the association between the dimensions, as it seems intuitive that CG is more strongly associated with AD than with PD.

Contraction and scaling

The contraction approach described here bears strong mathematical resemblance to the SFA: when estimating the disutility coefficients in SFA, the expression for $\delta^*(Q)$ as presented in Eq. 2 is multiplied by parameter λ while also plugging in for the α coefficients for CDs the values estimated for the core EQ-5D-5L descriptive system (i.e., estimated for $\delta(Q)$ as in Eq. 1).

In spite of the similarity of mathematical specifications, the motivation behind both approaches, SFA and the contraction introduced in this paper, is different. In the SFA, scaling allows coefficients from the core value set to be reused. This reduces the number of parameters that must be estimated in the BO valuation study. At the same time, it allows the importance of the CDs to change when the BO is appended to the health-state description. An additional benefit is to enforce a form of consistency between the value sets (core and BO) by ensuring that the relative importance of CDs and of levels within each CD is retained. Meanwhile, for the contraction approach the motivation relates to the attempt to remove the double counting of HRQoL benefits by reducing the disutilities of all dimensions to the extent that reflects the association between CDs and BO. Accordingly, the reduction in CD coefficients observed in an SFA valuation study may mitigate double counting, but it does not by itself demonstrate that the resulting six-dimensional value set is consistent with the existing core value set in the sense advocated in the current paper. Such consistency additionally depends on the joint distribution of CD and BO responses in measurement data. Importantly, the association observed in measurement data need not correspond to the reduction in CD disutilities elicited in preference studies. For instance, two dimensions may be correlated because they share a common cause, without necessarily being perceived as overlapping by respondents in valuation tasks.

In empirical studies using SFA, λ takes values in the range 0.7–0.9, which is in the same ballpark as the contraction parameter, κ, obtained in this paper. This similarity is a convenient coincidence: the reduction of disutilities of CDs resulting in preference studies from appending a BO to the descriptive system is similar in size to what is needed to remove the effect of double counting between CD and BO in measurement. This similarity perhaps is caused by respondents in valuation studies intuitively perceiving the information overlap between CDs and BOs. The consequence of this similarity is that when λ is retained in the production of a BO value set in SFA, only weak further modifications may be needed to ensure consistency of core and BO value sets (notably, when λ is substantially lower than 1, then an expansion, rather than contraction, may be needed to ensure consistency).

It is important to stress that SFA and the contraction approach proposed in this paper can be used jointly or separately. In particular, the contraction proposed here can be applied to a BO value set obtained with any valuation method.

Limitations

We see the following limitations of our study. First, one could argue that the proposed method modifies the elicited preferences. From this perspective, EQ-5D-5L+BO

preferences should instead be used at face value, including any scaling by $\lambda < 1$ estimated using the SFA. We see the appeal of this stance. However, our approach is based on more pragmatic views on the use of BOs in CUA. In this context, we think of EQ-5D-5L+BO descriptive systems not as independent instruments, but as extensions to the core EQ-5D-5L descriptive system that are made available to increase the measurement sensitivity. In consequence, it is paramount that the differences in CUA results are clearly interpretable and can be directly linked specifically to the increased sensitivity.

Accordingly, our approach does not address possible informational overlap among the five CDs themselves. Indeed, the importance of the remaining CDs may change when one of the five dimensions is removed, as demonstrated for self-care (Yang et al. 2024). Such overlap may exist, but it is part of the established EQ-5D-5L descriptive and valuation framework that serves as the benchmark in the present analysis. Extending the contraction principle to redesign or recalibrate the core instrument would be a distinct and substantially broader research question.

Second, even if one accepts the need to adjust estimated disutilities, not all possible consistency criteria can be satisfied simultaneously. The choice of criterion is therefore necessary and is likely to remain normative. As noted above, making the two value sets consistent with respect to the standard deviation of the conditional mean implies that life-extending technologies will generate fewer additional QALYs under the BO value set than under the core value set. Obviously, using a non-contracted BO value sets would produce even fewer QALYs. The fact that accounting for health problems makes the value of added life-years smaller is just a property of the QALY approach [subject to research, see Basu et al. 2020], and accounting for more health problems with BOs makes this issue more pronounced.

Third, the approach we propose in the study uses only association between CDs and BO and it ignores altogether the mechanism which produces the association, for instance, the direction of the causal link. The causal direction may differ across BOs. For example, impaired mobility may restrict social relationships, in which case a CD affects the BO. Conversely, vision problems may reduce the ability to perform usual activities, so the BO affects a CD. In other cases, a third factor, such as age, may affect both the CDs and the BO. Future research is recommended to discuss whether consistency criterion should reflect the possibility of varying mechanisms, and the following example suggests that it should. Suppose that a CD and a BO are associated only because both worsen with age. In this case, one could argue that the two dimensions do not contain overlapping information. The BO may provide distinct information despite the observed association, and contraction may therefore be unjustified. Future studies should extend the method proposed in this paper to account for selected demographic covariates (age in particular).

Additionally, the conditional association between dimension levels in cross-sectional HRQoL data does not necessarily translate into an equivalent association between within-person changes over time. Cross-sectional data have a direct interpretation for changes from an observed state to full health, (11111+1), because baseline dimension levels determine the improvements required to reach that target. For partial or heterogeneous treatment effects, however, longitudinal data may provide a more appropriate basis for defining consistency. Future studies could extend the proposed approach by using the conditional distribution of BO changes given changes in the CDs.

Fourth, in spite of the similarity of results for the two population types obtained in the paper, substantial differences may occur in the future for other BOs or when additional data are collected with EQ-5D Bolt-on Toolbox. Hence, a decision is needed on which population is the appropriate source of measurement data, and the present paper does not offer a conclusive answer. There are arguments for using general-population data. Such data are easier to collect in large samples, and the target population is more clearly defined. By contrast, a BO may be relevant to several different patient populations. The use of general-population data can also be justified by the motivation for consistency. Larger QALY gains obtained with a BO may be expected in the patient population for which the BO was developed, but similar gains may be more difficult to justify outside that population. There are also arguments for using patient data. Patient samples usually contain more observations at severe levels and may therefore provide more information about the association between dimensions. They also correspond more closely to the populations in which the BO is expected to be used.

Another limitation concerns the use of disutility coefficients originating from different countries. Ideally the CD and BO coefficients would be derived from valuation studies conducted in the same country and target population. The contraction parameter depends on their relative magnitudes, which may vary across national valuation studies. Combining coefficients from different settings may therefore produce a different value of κ than would be obtained using a fully country-specific pair of core and BO value sets. Our sensitivity analysis using the Polish core value set suggests that the choice of core value set may systematically affect the magnitude of contraction, while having little impact on the relative differences between BOs (see Supplementary Materials). In the present methodological illustration, we used a common UK core value set to facilitate comparisons

across BOs and because five of the nine BO coefficient sets were derived from UK studies. Accordingly, the reported contraction parameters should not be interpreted as definitive country-specific adjustments. In an applied development of a BO value set, both sets of coefficients should preferably be derived from the same target population.

Additionally, while differentiating the contraction by CD may seem sensible and justified to reflect different correlation between a given BO and individual CDs, the approach we proposed in Section "Alternative approaches to calculating κs" is rather ad hoc and lacks a strong theoretical foundation. Perhaps a better founded approach could be devised in future methodological work.

Finally, we obviously have to acknowledge limited data in patient population for some BOs, especially for CG and SI, and no data altogether for HE.

Possible extensions

The main approach does not depend on the specific functional forms of $\delta(\cdot)$ and $\delta^*(\cdot)$. The two functions need not have the same form: for example, one or both may contain interaction terms. The method requires only that $\delta^*(\cdot)$ can be multiplied by a constant. This flexibility does not apply to the alternative approach in which contraction differs across individual CDs. That approach requires $\delta^*(\cdot)$ to contain no interactions between the CDs, although such interactions may be present in $\delta(\cdot)$. In particular, as mentioned above, our approach may be combined with SFA. It is worth noticing that our method can be easily extended to multiple BOs included simultaneously, or a BO having multiple items.

We also believe that an approach similar to ours could be devised to define and ensure consistency of value sets in other contexts. First, the EQ-5D family is increasingly used in younger populations. The EQ-5D-Y-3L and EQ-5D-Y-5L were designed for children and adolescents (Devlin et al. 2022; Pan et al. 2025), whereas the EQ-TIPS is being developed for infants and toddlers (Verstraete et al. 2020). Methods similar to ours could be tested to measure the consistency of value sets along the life cycle. Second, as noted in the Introduction, various preference-based measures are used in CUA. The approach proposed in this paper could likely be adapted to studying consistency between various such measures and associated value sets.

Conclusion

BOs are useful in some clinical conditions to increase EQ-5D-5L sensitivity in measuring HRQoL. In CUA, BO value sets may complement the core EQ-5D-5L value sets. To increase acceptance of BOs, it is important to indicate whether the impact on CUA results is justified by the increased sensitivity or whether it also reflects double-counting of HRQoL by CDs and BO. We propose a definition of consistency between the two types of value sets to help make this distinction, and we propose a method that can be included in development of a BO value set that adheres to this consistency. For the tested BOs, a small to moderate contraction of disutility values is needed to make the BO value sets consistent.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10198-026-02009-z.

Acknowledgements The authors thank the EuroQol Research Foundation for funding the present study. However, the views presented here may not be shared by this foundation or the EuroQol Group. The authors thank those who shared their measurement data. The authors thank the EQ-DAPHNIE project for sharing their data. That project is supported by operating grants from the EuroQol Research Foundation (reference 367-RA). We thank Sarah Dewilde, Bas Janssen, Fredrick Purba, Mercédesz Angyal, Lajos Vince Kemény, and Eszter Mercédesz Müller. We thank the University of Birmingham, UK for granting access to data from The Birmingham COPD Cohort Study, which was funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0109-10061). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. The Birmingham COPD Cohort study is part of The Birmingham Lung Improvement StudieS – BLISS. The authors also thank Donna Rowen for sharing information on UK valuation results, Zhihao Yang for sharing unpublished data on bolt-on disutility, and Eleanor Pullenayegum for useful suggestions to an earlier version of the paper.

Author Contributions Substantial contributions to the conception or design of the work: all authors. Data analysis: MJ. Interpretation of data for the work: all authors. Drafting the work or reviewing it critically for important intellectual content: all authors. Final approval of the version to be published: all authors.

Funding The study was funded by the EuroQol Research Foundation (2009-RA).

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

Declarations

Competing Interests The study was funded by the EuroQol Research Foundation (2009-RA). However, the views presented here may not be shared by this foundation or the EuroQol Group. All authors are members of the EuroQol Group. All authors report receiving grants from the EuroQol Research Foundation and funding for travelling to and attending meetings and conferences organized by the EuroQol Research Foundation. Dr Rencz is employed by the EuroQol Research Foundation.

Ethics Approval Because no new data were collected for this study, no ethical approval was deemed necessary.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Adab, P., D. A. Fitzmaurice, A. P. Dickens, J. G. Ayres, H. Buni, B. G. Cooper, A. J. Daley, A. Enocson, S. Greenfield, K. Jolly, S. Jowett, K. Kalirai, J. L. Marsh, M. R. Miller, R. D. Riley, W. S. Siebert, R. A. Stockley, A. M. Turner, K. K. Cheng, and R. E. Jordan. 2017. Cohort profile: The Birmingham Chronic Obstructive Pulmonary Disease (COPD) cohort study. *International Journal of Epidemiology* 46:23. https://doi.org/10.1093/ije/dyv350.

Angyal, M. M., M. F. Janssen, P. L. Lakatos, V. Brodszky, and F. Rencz. 2025. The added value of the cognition, dining, gastrointestinal problems, sleep and tiredness bolt-on dimensions to the EQ-5D-5L in patients with coeliac disease. *The European Journal of Health Economics* 26:473–485. https://doi.org/10.1007/s10198-024-01719-6.

Roudijk, B. K Ludwig, and N Devlin. *EQ-5D-5L value set summaries*, 55–212. Springer International Publishing, Cham, 2022. https://doi.org/10.1007/978-3-030-89289-0_4

Basu, A., J. Carlson, and D. Veenstra. 2020. Health years in total: A new health objective function for cost-effectiveness analysis. *Value in Health* 23:96–103.

Buchholz, I., M. F. Janssen, T. Kohlmann, and Y.-S. Feng. 2018. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics* 36:645–661.

Davis, J. C., T. Liu-Ambrose, K. M. Khan, M. C. Robertson, and C. A. Marra. 2012. SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. *Osteoporosis International* 23:1849–1857.

Devlin, Nancy. David Parkin, and Bas Janssen. 2020. *Methods for Analysing and Reporting EQ-5D Data*. Springer Cham, 1 edition, https://doi.org/10.1007/978-3-030-47622-9.

Devlin, N., T. Pan, S. Kreimeier, J. Verstraete, E. Stolk, K. Rand, and M. Herdman. 2022. Valuing EQ-5D-Y: the current state of play. *Health and Quality of Life Outcomes* 20:105.

Dewilde, S., G. Phillips, S. Paci, F. De Ruyck, N. H. Tollenaar, and M. F. Janssen. 2024. The burden patients with myasthenia gravis experience in terms of breathing, fatigue, sleep, mental health, discomfort and usual activities in comparison to the general population. *Advances in Therapy* 41:271–291. https://doi.org/10.1007/s12325-023-02704-w.

Dewilde, S., M. F. Janssen, G. Phillips, B. Mulhern, and F. Rencz. 2025. Comparing psychometric properties of six 5-level and 3-level EQ-5D Bolt-Ons in a large, multinational, longitudinal general population sample. *Value in Health* S1098–3015:02566–5. https://doi.org/10.1016/j.jval.2025.09.009.

EuroQol. 2026. EQ-5D Bolt-on Toolbox. https://euroqol.org/research-at-euroqol/eq-bolt-ons/ [Accessed: 16 Feb 2026].

Feng, Y.-S., T. Kohlmann, M. F. Janssen, and I. Buchholz. 2021. Psychometric properties of the EQ-5D-5L: A systematic review of the literature. *Quality of Life Research* 30:647–673.

Geraerds, A. J. L. M., G. J. Bonsel, M. F. Janssen, A. P. Finch, S. Polinder, and J. A. Haagsma. 2021. Methods used to identify, test, and assess impact on preferences of Bolt-Ons: A systematic review. *Value in Health* 24:901–916.

Golicki, D., M. Jakubczyk, K. Graczyk, and M. Niewada. 2019. Valuation of EQ-5D-5L health states in Poland: The first EQ-VT-based study in central and Eastern Europe. *Pharmacoeconomics* 37:1165–1176.

Grutters, J. P., M. A. Joore, F. van der Horst, H. Verschuure, W. A. Dreschler, and L. J. Anteunis. 2007. Choosing between measures: Comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. *Quality of Life Research* 16:1439–1449.

Hoogendoorn, M., M. Oppe, M. R. S. Boland, L. M. A. Goossens, E. A. Stolk, and M. P. M. H. Rutten-van Mölken. 2019. Exploring the impact of adding a respiratory dimension to the EQ-5D-5L. *Medical Decision Making* 39:393–404.

Hoogendoorn, M., S. Jowett, A. P. Dickens, R. Jordan, A. Enocson, P. Adab, M. Versteegh, and M. R. M ölken. 2021. Performance of the EQ-5D-5L plus respiratory Bolt-On in the birmingham chronic obstructive pulmonary disease cohort study. *Value in Health* 24:1667–1675.

Jakubczyk, M. 2023. What if 0 is not equal to 0? Inter-personal health utilities anchoring using the largest health gains. *The European Journal of Health Economics* 24:1217–1233.

Johnson, JA. MF Janssen, F Al Sayah, H Bailey, M Gandhi, D Golicki, N Gutacker, E Lubetkin, B Mulhern, FD Purba, JM Ramos-Goñi, D Scott, H Short, T Sullivan, R Viney, Z Yang, and V Zárate. 2025. EuroQol data for assessment of population health needs and instrument evaluation (EQ-DAPHNIE): a study for enhancing population health assessment. *Quality of Life Research* 34:3321–3334. https://doi.org/10.1007/s11136-025-03983-2

Kennedy-Martin, M., B. Slaap, M. Herdman, M. van Reenen, T. Kennedy-Martin, W. Greiner, J. Busschbach, and K. S. Boye. 2020. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *The European Journal of Health Economics* 21:1245–1257.

Kharroubi, S. A., Y. S. Beyh, J. Brazier, and Y. Yang. 2020. Modelling a preference-based index for EQ-5D-3L and EQ-5D-3L + Sleep using a Bayesian framework. *Quality of Life Research* 29:1495–1507.

Kim, H., G. Cook, S. Goodall, and D. Liew. 2021. Comparison of EQ-5D-3L with QLU-C10D in metastatic melanoma using cost-utility analysis. *Pharmacoeconomics Open* 5:459–467.

Ludwig, K., J.-M. Graf von der Schulenburg, and W Greiner. 2018. German value set for the EQ-5D-5L. *Pharmacoeconomics* 36:663–674.

Mulhern, B. J., C. Sampson, P. Haywood, R. Addo, K. Page, D. Mott, K. Shah, M. F. Janssen, and M. Herdman. 2022. Criteria for developing, assessing and selecting candidate EQ-5D bolt-ons. *Quality of Life Research* 31:3041–3048.

Müller, E. M., A. Nikl, M. Krebs, P. Holló, V. Brodszky, L. V. Kemény, and F. Rencz. 2025. Psychometric benefits of adding bolt-ons to the EQ-5D-5L in populations undergoing minimally invasive cosmetic procedures. The European Journal of Health Economics 26:1233-1247. https://doi.org/10.1007/s10198-025-01772-9

NICE. NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE. 2025. *Draft guidance consultation. Dupilumab for treating severe chronic rhinosinusitis with nasal polyps*. https://www.nice.org.uk/guidance/GID-TA11630/documents/consultation-document. Accessed: 31 July 2025.

Norman, R., B. Mulhern, E. Lancsar, P. Lorgelly, J. Ratcliffe, D. Street, and R. Viney. 2023. 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* 41:427–438.

Pan, T., J. M. Ramos-Goñi, B. Roudijk, S. Xie, F. Xie, Z. Yang, B. Mulhern, R. Norman, and N. Devlin. 2025. Testing the valuation of the EQ-5D-Y-5L in adults and adolescents: Results from a 5-country study and implications for the descriptive system. *Value in Health* 28:1900–1910.

Pangestu, S. F Purba, H Setyowibowo, C Mukuria, and F Rencz. 2025. comparative psychometric analysis of eq-5d-5l-format bolt-ons and eq-hwb bolt-on items in patients with breast cancer [Poster presented at the 5th EuroQol Early Career Researchers Meeting, Barcelona, Spain].

Payakachat, N., M. M. Ali, and J. M. Tilford. 2015. Can the EQ-5D detect meaningful change? A systematic review. *Pharmacoeconomics* 33:1137–1154.

Pickard, A. S., E. H. Law, R. Jiang, E. Pullenayegum, J. W. Shaw, F. Xie, M. Oppe, K. S. Boye, R. H. Chapman, C. L. Gong, A. Balch, and J. J. V. Busschbach. 2019. United States valuation of EQ-5D-5L health states using an international protocol. *Value in Health* 22:931–941.

Rencz, F., and M. F. Janssen. 2024. Testing the psychometric properties of 9 Bolt-Ons for the EQ-5D-5L in a general population sample. *Value in Health* 27 (7): 943–954.

Rencz, F., V. Brodszky, L. Gulácsi, D. Golicki, G. Ruzsa, A. S. Pickard, E. H. Law, and M. Péntek. 2020. Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value in Health* 23 (9): 1235–1245.

Rencz, F., S. Pangestu, B. Mulhern, A. P. Finch, and M. F. Janssen. 2025. Development and use of cognition Bolt-Ons for the EQ-5D-3L and EQ-5D-5L: A systematic review. *Value in Health* 28 (11): 1752–1769.

Rencz, F. B Slaap, N Devlin, E Stolk, B Mulhern, and the International Bolt-on Advisory Group. 2026. EQ-5D Bolt-on Toolbox: concept, development, valuation, applications and future directions. *Value in Health*. https://doi.org/10.1186/s12955-026-02635-4.

Rowen, D., C. Mukuria, N. Bray, J. Carlton, L. Longworth, D. Meads, Y. Oluboyede, C. O'Neill, and Y. Yang. 2026. A UK value set for the EQ-5D-5L. *Value in Health*. https://doi.org/10.1016/j.jval.2026.03.008.

Swinburn, P., A. Lloyd, K. S. Boye, E. Edson-Heredia, L. Bowman, and B. Janssen. 2013. Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: Lessons learned from a feasibility study in the UK. *Value in Health* 16:1156–1162.

Versteegh, M. M., K. M. Vermeulen, S. M. A. A. Evers, G. A. de Wit, R. Prenger, and E. A. Stolk. 2016. Dutch tariff for the five-level version of EQ-5D. *Value in Health* 19:343–352.

Verstraete, J., L. Ramma, and J. Jelsma. 2020. Validity and reliability testing of the Toddler and Infant (TANDI) health related quality of life instrument for very young children. *Journal of Patient-reported Outcomes* 4:94.

Whitehead, S. J., and S. Ali. 2010. Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin* 96:5–21.

Yang, Y., D. Rowen, J. Brazier, A. Tsuchiya, T. Young, and L. Longworth. 2015. An exploratory study to test the impact on three ''Bolt-On'' items to the EQ-5D. *Value in Health* 18:52–60.

Yang, Z., K. Rand, J. Busschbach, and N. Luo. 2023. Cross-attribute level effects models for modeling modified 5-level version of EQ-5D health state values: Is less still more? *Value in Health* 26:865–872.

Yang, Z., K. Rand, and N. Luo. 2024. Valuation of multidimensional health states with a Bolt-On: Is there a shortcut? *Value in Health* 27:1544–1552.

Yang, Z., K. Rand, A. Finch, and N. Luo. 2026. Valuation of EQ-5D-5L with 2 Bolt-On items: Further evaluation of the scaling factor model. Value in Health 29(8):1443-1451 . https://doi.org/10.1016/j.jval.2026.02.012

Publisher"s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.