PharmacoEconomics - Open (2026) 10:503–518
https://doi.org/10.1007/s41669-026-00645-y

ORIGINAL RESEARCH ARTICLE

Exploring the Sensitivity of Utilities Predicted by the EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D Multi-Attribute Utility Instruments to SF-36 Dimensions

Josephine N. A. Tetteh[1] · **Michael Schlander**[1,2,3]

Received: 16 July 2025 / Accepted: 8 March 2026 / Published online: 13 April 2026

Abstract

Background Multi-attribute utility instruments (MAUIs) are commonly used in health economics to measure health-related quality of life (HRQoL), yet their sensitivity to different health domains varies. This study examines the sensitivity of six widely used MAUIs—EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D—to the eight dimensions of the SF-36 survey.

Methods We analyzed the associations between SF-36 dimensions and utility scores generated by each MAUI using regression models, focusing on the eight SF-36 domains as predictors. Our analyses used data from the Multi-Instrument Comparison (MIC) project, a cross-national project comprising 8022 respondents from Australia, Canada, Germany, Norway, the United Kingdom, and the United States, including both general population participants and individuals with a range of chronic health conditions. The sensitivity of each instrument was further evaluated through simulations of health interventions targeting mental health, pain relief, stress management, and post-surgical recovery.

Results The analysis revealed distinct sensitivity patterns across instruments. The SF-6D and 15D were particularly sensitive to social and role-oriented domains, while EQ-5D and HUI3 demonstrated greater sensitivity to physical health dimensions, especially bodily pain and physical functioning. AQoL-4D and AQoL-8D showed strong sensitivity to mental health, indicating their suitability for mental health-focused interventions. Regression models identified mental health, bodily pain, and physical functioning as the primary predictors of utility scores across MAUIs. Simulation results further highlighted that mental health interventions were best captured by AQoL-8D and HUI3, while EQ-5D and HUI3 were most sensitive to pain-focused interventions.

Conclusions The choice of MAUI significantly impacts observed HRQoL outcomes depending on the health domains most relevant to the intervention. This study underscores the need for careful instrument selection in HRQoL assessments to ensure alignment with specific health contexts. Future research should validate these sensitivity patterns across diverse populations and clinical applications to optimize MAUI selection in health economic evaluations.

1 Introduction

Health-related quality of life (HRQoL) is a critical outcome measure in health economics, clinical research, and healthcare decision making. HRQoL evaluations provide insights into how individuals perceive their health status across multiple dimensions, including physical, mental, and social well-being [1–5]. Multi-attribute utility instruments (MAUIs) are widely used in this context to quantify health outcomes as single utility values [6]. Instruments like the EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D have become instrumental in summarizing HRQoL for economic evaluations, enabling comparisons across diverse health interventions and populations [7]. These instruments are well established tools that allow researchers and policymakers to evaluate the impact of health interventions and inform resource allocation through cost-utility analyses.

Each MAUI has a unique structure, item set, and weighting scheme, which influences its sensitivity to various aspects of health and may have practical implications. Research indicates that some instruments may be better

✉ Michael Schlander
ms@michaelschlander.com

1 Division of Health Economics, German Cancer Research Center (DKFZ), Heidelberg, Germany

2 Institute for Innovation and Valuation in Health Care (InnoVal-HC), Wiesbaden, Germany

3 Faculty of Economics and Social Sciences, Alfred Weber Institute, University of Heidelberg, Heidelberg, Germany

Key Points for Decision Makers

There is growing use of multi-attribute utility instruments (MAUIs) to assess health-related quality of life (HRQoL) in health economic evaluations, but these instruments differ in the aspects of health they best capture. This study provides a direct comparison of six widely used MAUIs—EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D—based on their sensitivity to eight health domains from the SF-36 survey.

Statistical analyses and simulations revealed that EQ-5D and HUI3 are particularly responsive to physical health improvements, such as pain and mobility. In contrast, AQoL-4D and AQoL-8D are more sensitive to mental health changes, while SF-6D and 15D are better at detecting changes in social and role functioning.

The choice of utility instrument can significantly affect the outcomes of health evaluations. Decision makers should select MAUIs that align closely with the expected health outcomes of an intervention to ensure accurate measurement of benefits. Further research is needed to validate these findings across diverse clinical settings.

suited for certain populations or conditions, depending on their dimensional coverage and sensitivity profiles. For instance, the EQ-5D, developed by the EuroQol Group, is concise and widely applicable, making it one of the most extensively used instruments in clinical trials and health surveys [8–10]. However, it has been critiqued for its limited sensitivity in capturing mental health and social functioning, areas where more complex instruments like AQoL-8D or SF-6D may offer a broader perspective [11–14]. Furthermore, HUI3 has been found to be highly sensitive to physical functioning and mobility-related issues, making it valuable in studies focusing on conditions affecting mobility [15]. The diverse structural and methodological approaches underlying these instruments underscore the need for a thorough understanding of how each MAUI performs in capturing various HRQoL dimensions, particularly when selecting an appropriate tool for a specific study or intervention.

The SF-36 health survey is among the most widely used generic HRQoL profiles and assesses health across eight key dimensions: Physical Functioning (PF), Role Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role Emotional (RE), and Mental Health (MH) [16, 17]. It is frequently used as a reference measure for evaluating the sensitivity and validity of other HRQoL instruments due to its extensive validation, reliability, and comprehensive health coverage. Several studies have used SF-36 dimensions as predictive and comparative variables in psychometric evaluations of MAUIs to gain a deeper understanding of each instrument's responsiveness to particular health domains.

In the context of MAUIs, sensitivity is a measure of how well an instrument reflects variations in specific aspects of health, such as physical, mental, or social well-being [18]. Each instrument places differing emphasis on health dimensions, which can lead to variations in results based on the population and health condition being assessed. Numerous studies have investigated the sensitivity of individual MAUIs, with findings showing substantial variation in sensitivity across instruments and patient populations [7, 12, 18, 19].

Previous studies have shown that although the SF-6D and EQ-5D are frequently used in general population health studies, they prioritize different health dimensions and produce distinct results across various patient populations and conditions. The SF-6D, typically valued using the Standard Gamble (SG) method, often reports higher scores among less healthy individuals due to its valuation approach rather than the instrument itself. Structurally, it emphasizes social functioning and role limitations, whereas the EQ-5D, usually valued using the Time Trade-Off (TTO) method, is more sensitive to physical functioning and offers a broader scoring range across severity levels [20]. Other research has highlighted that HUI3 is particularly effective in identifying changes in physical functioning and hearing-related issues due to its specific dimension items [21, 22], whereas AQoL instruments, designed with extensive psychosocial items, may be better at capturing well-being in mental health conditions [23]. Furthermore, the 15D, which has multiple response options across its 15 dimensions, has been praised for its nuanced sensitivity particularly in capturing moderate health states [11].

Differences in the sensitivity of MAUIs have important implications for healthcare decision making and health technology assessment (HTA). Utility values derived from these instruments are the foundation of quality-adjusted life years (QALYs), which are used to evaluate cost effectiveness and guide reimbursement and resource allocation decisions. When an instrument is more or less sensitive to particular health domains, it may systematically over- or under-estimate changes in HRQoL, leading to differences in estimated QALY gains for the same intervention. Previous research has demonstrated that such discrepancies can materially affect incremental cost-effectiveness ratios (ICERs) and, consequently, the conclusions of economic evaluations [8, 12, 24]. For instance, instruments emphasizing physical health (such as EQ-5D) may favor interventions targeting mobility or pain, while those with stronger psychosocial coverage (such as AQoL-8D or SF-6D) may capture greater benefits for mental health interventions. This highlights the importance of aligning instrument choice with the nature of health changes being assessed. The present study adds to this literature by providing a systematic, dimension-level comparison of six major MAUIs using SF-36 domains as predictors. By identifying which SF-36 dimensions most strongly influence each MAUI's utility scores, our findings help clarify how instrument sensitivity translates into potential differences in cost-effectiveness outcomes and inform the selection of appropriate MAUIs for policy and HTA applications.

While previous work, including the influential Multi-Instrument Comparison (MIC) studies by Richardson and colleagues [12, 14, 19], has explored the sensitivity and construct validity of several MAUIs, these analyses have generally focused on a subset of instruments or a limited range of health dimensions. This study extends that foundational work by systematically comparing six MAUIs across all eight SF-36 dimensions using the same dataset, thereby providing a broader and more integrated understanding of their relative sensitivity profiles. In addition to building upon this earlier evidence, this study seeks to explicitly link the concept of sensitivity to its implications for health technology assessment (HTA) and cost-effectiveness analysis (CEA). Differences in MAUI sensitivity can translate into variations in predicted utilities and thus influence the estimation of quality-adjusted life years (QALYs). A clearer understanding of these differences is essential to ensure that health policy decisions and resource allocation are based on instruments that reflect the most relevant dimensions of health for a given intervention or population.

Moreover, inconsistencies in study design, sample characteristics, and statistical approaches across sensitivity studies complicate the ability to draw firm conclusions on which MAUIs offer the most comprehensive sensitivity profile. Therefore, a broader and more systematic comparison of MAUIs using SF-36 dimensions as predictors would offer valuable information on each instrument's sensitivity to physical, mental, and social health changes.

This study aims to address this gap by evaluating the sensitivity of six widely used MAUIs—EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D—to individual SF-36 dimensions across a large, cross-sectional sample, focusing on how each instrument's unique structure and content influence its capacity to capture changes across different health domains. The study has three main objectives: (1) to determine the unique contributions of individual SF-36 dimensions to observed MAUIs, (2) to identify which SF-36 dimensions are most significant in determining overall quality of life (QoL) as measured by various MAUIs and evaluate the collective predictive power of these dimensions on QoL utility scores, and (3) to analyze the sensitivity of each MAUI to changes in specific SF-36 dimensions to identify which instrument is more sensitive to changes in a particular dimension.

By examining these differences, this research provides both methodological and applied insights: methodologically, it highlights the conceptual and empirical distinctions among widely used MAUIs; and practically, it discusses how these differences may affect cost-effectiveness outcomes and healthcare decision making. Specifically, understanding how each MAUI responds to changes in SF-36 domains can help decision makers anticipate potential discrepancies in QALY estimation and choose instruments that align with the intervention's targeted health outcomes. In doing so, this study contributes to the broader understanding of the relative advantages and limitations of these MAUIs in capturing HRQoL across diverse health domains.

2 Methods

This study utilized data from the MIC project, which was conducted in the early 2010s and involved a sample drawn from six countries: Australia, Canada, Germany, Norway, the United Kingdom, and the United States [12]. The MIC project's primary objective was to evaluate and compare a range of health measurement instruments, with recruitment managed through an online survey distributed by CINT Pty Ltd., a professional online panel company. To achieve a representative demographic sample, CINT selected respondents from its database, ensuring diversity in age, gender, and health conditions. For accessibility, the survey was professionally translated into German and Norwegian and validated by local experts to ensure cultural and linguistic accuracy.

Participants in the MIC survey completed an extensive set of HRQoL instruments, including the following six primary multi-attribute utility instruments (MAUIs): EQ-5D-5L, HUI3, SF-6D, 15D, AQoL-4D, and AQoL-8D. Additional instruments included the generic HRQoL profile SF-36, the Quality of Wellbeing (QWB) scale, three subjective well-being scales, and three other HRQoL instruments. The initial sample consisted of 8022 respondents. After thorough data cleaning and quality checks, the final sample size was adjusted to 6845 participants.

2.1 Instrument Description

AQoL-4D: The AQoL-4D was designed as the foundational multi-attribute utility instrument of the AQoL series to assess health-related quality of life across physical, psychological, and social dimensions. With 12 items distributed across four dimensions—independent living, social relationships, physical senses, and psychological well-being—the instrument employs a multiplicative utility formula to combine items and dimensions. The AQoL-4D was specifically developed to achieve psychometric orthogonality between dimensions, distinguishing it from later AQoL versions which adopted econometric corrections. The instrument's preference weights were derived using the TTO valuation method. The AQoL-4D has been extensively validated and translated into multiple languages, allowing flexibility in research applications [23].

AQoL-8D: The AQoL-8D was designed to assess health states affected by both physical and mental conditions. It is part of the AQoL series, which aims to gauge how health issues impact an individual's life by measuring the extent of disability caused by illness. With 35 items spread across eight dimensions, the AQoL-8D was developed using psychometric methods to ensure its content validity. Three dimensions—namely independent living, pain, and senses—combine to form a physical super-dimension, while the remaining five—mental health, happiness, coping, relationships, and self-worth—comprise a mental super-dimension. This instrument is expansive, capable of delineating billions of health states. Its development involved 322 individuals with mental health conditions and 306 members of the general public, utilizing the TTO valuation method [25].

EQ-5D-5L: The EQ-5D-5L is an advanced version of the EQ-5D developed by the EuroQol Group, comprising five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. This tool was created to address common health domains that align with patients' core concerns, providing a comprehensive assessment of HRQoL. Originally, the EQ-5D included only three response levels, resulting in 243 potential health states. However, in 2009, the tool was upgraded to five response levels per item to enhance sensitivity, reliability, and reduce ceiling effects, thus expanding its capacity to measure 3125 distinct health states. The EQ-5D-5L was scored using valuation tariffs from multiple countries, specifically England, Canada, the US, and Germany, ensuring broad applicability in this international study. These tariffs, derived from TTO preference data, provide a robust basis for evaluating health states within diverse populations [10, 26].

SF-6D: The SF-6D in this study is derived from the SF-36, which was developed to measure general health status across eight dimensions, forming two summary scores for physical and mental health. The SF-6D consists of six multi-level dimensions (physical functioning, role limitation, social functioning, pain, mental health, and vitality) and describes 18,000 health states. Utility values were obtained using the standard gamble method to evaluate preferences for 249 health states with data from 611 UK participants. A non-parametric Bayesian approach, providing greater predictive power, was used for the utility algorithm in this study [27]. Although based on SF-36 items, the SF-6D applies population-based preference weights through a non-linear scoring procedure, producing utilities that differ conceptually and empirically from SF-36 scores. As only a subset of SF-36 items is included and certain domains (e.g., role limitations and bodily pain) are only partially represented, correlations between SF-6D utilities and SF-36 domain scores are expected to be moderate.

HUI3: The HUI3, a modified version of the HUI2, includes domains such as vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain, defining 972,000 health states between five and six levels. The HUI3 primarily focuses on "within-the-skin" symptoms, with few psychosocial domains. Utility weights were derived from a representative sample of adult Canadians using the visual analog scale (VAS) for valuation, with some health states also valued using the standard gamble technique to transform VAS values into standard gamble values [15].

15D: The 15D instrument includes 15 dimensions (mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity). Each dimension has four or five levels, allowing the 15D to define billions of health states. The instrument's valuation was based on five random samples of 500 people each from the Finnish general population, using a variant of the VAS [11].

SF-36: The Short Form 36 (SF-36) originated from the Rand Medical Outcomes Study's 245 items [28] and is now the most widely used instrument globally for measuring HRQoL. From the original 40 physical and mental health concepts, eight were selected using psychometric methods. The SF-36 has repeatedly demonstrated reliability across various countries and its validity and sensitivity have been confirmed in numerous settings and nations. The SF-36, or Short Form Health Survey, assesses HRQoL across eight domains: PF, RP, BP, GH, VT, SF, RE, and MH. It also provides two summary scores: the Physical Component Summary (PCS) and the Mental Component Summary (MCS). The SF-36 is widely used and validated, serving as a standard tool in health surveys for comparing health status across different diseases, treatments, and populations [29].

In this study, the scores for the eight dimensions of the SF-36 are adjusted to a scale of 0 to 1 for simpler interpretation of the findings. Table 1 highlights the substantial differences in the size and content of the MAU instruments, which are described in detail in Brazier [24] and Richardson et al. [12].

Table 2 summarizes the country-specific value sets and valuation techniques for each MAUI included in this study.

2.2 Analysis

The first objective of this study was to identify the contributions of each single SF-36 dimension to variations in MAUIs. Pearson correlation coefficients were used to assess the strength and direction of the linear relationship between the MAUIs and SF-36 dimensions. A correlation >0.5 was considered a strong association, between 0.3 and 0.49 a moderate association and <0.3 weak. To model this relationship, simple linear regressions with MAUIs as the dependent variable and the individual SF-36 dimension scores as independent variables were used. Beta coefficients are then examined, and their statistical significance are assessed to quantify the impact of each SF-36 dimension on MAUI. The model formulation is as follows:

$$MAUI = \beta_0 + \beta_i X_i + \epsilon \quad (1)$$

where MAUI is EQ-5D, HUI3, SF-6D, 15D, AQoL-4D, and AQoL-8D utility values, β_0 is a constant, X_i is an individual SF-36 dimension, β_i is a coefficient denoting the contribution of SF-36 dimension X_i to the MAUI score, and ϵ accounts for the error term capturing unexplained variability.

The second objective was to determine which dimensions of the SF-36 are most influential in determining MAUIs and to evaluate their predictive power. To assess the collective influence of all SF-36 dimensions on changes in quality of life utility scores, a multiple linear regression model was used. This approach examines how variations in all eight dimensions simultaneously predict changes in utility values. The model equation is:

$$MAUI = \beta_0 + \sum_{i=1}^{8} \beta_i X_i + \in \quad (2)$$

where MAUI represents the MAUI scores, $(X_1, X_2, \ldots, X_8)$ denote the SF-36 dimensions, β_0 stands for the intercept (constant), β_i is the coefficients indicating the contributions of each SF-36 dimension to the quality of life utility value, and ϵ captures the error term representing unexplained variability. Since there are eight dimensions, $i = 1to8$.

To determine which variable is more influential in the multiple linear regression model in equation 2, we consider the magnitudes of the estimated β_i coefficients. The

Table 2 Overview of value sets, valuation methods, and references for the six MAUIs examined

Instrument	Country	Valuation method	References
EQ-5D-5L	UK	TTO	Van Hout et al. (2012) [20]
SF-6D	UK	SG	Brazier et al. (2002) [27]
HUI3	Canada	SG	Feeny (2002) [21]
15D	Finland	VAS	Sintonen (2001) [11]
AQoL-4D	Australia	TTO	Hawthorne et al. (2001) [14]
AQoL-8D	Australia	TTO	Richardson (2014) [19]

MAUIs multi-attribute utility instruments, *SG* standard gamble, *TTO* time trade-off, *VAS* visual analog scale

Table 1 Content of the six multi-attribute utility instruments (number of items) adapted from [24] and [12]

Dimension	EQ-5D	SF-6D	HUI3	15D	AQOL-4D	AQoL-8D
Physical						
Coping/control/physical ability		1		1		2
Bodily function/self-care	1			3	2	1
Dexterity			1			
Pain/discomfort	1	1	1	1	1	2
Senses (vision/hearing)			2	2	2	2
Usual activities/work function	1	1		1	1	4
Mobility/walking	1		1	1		1
Communication due to speech problems			1	1		1
Vitality						
Psycho-social						
Sleeping				1		1
Well-being/depression/anxiety/happiness	1	1	1	3	3	7
General satisfaction						4
Self-esteem/identity						2
Cognition/memory			1	1		
Relationships/social function		1			3	6
Family		1				1
Intimacy (including sexual relations)				1		1
Total number of items	**5**	**6**	**8**	**15**	**12**	**35**

coefficients represent the average change in the dependent variable (MAUI score) for a one-unit change in the independent variable (SF-36 dimensions), assuming all other variables are held constant. A positive coefficient suggests that higher scores on that dimension correspond to higher scores on the MAUI, while a negative coefficient indicates the opposite relationship. The magnitude of the coefficient reflects the strength of this relationship. If the coefficient is zero, it signifies that the model does not detect any association.

The third objective was to determine the sensitivity of each MAUI to changes in specific SF-36 dimensions. To achieve this, we utilized the same multiple regression models from equation 2 focusing on the magnitudes of the coefficients for each individual SF-36 dimension across the various MAUIs. This comparative analysis allowed us to determine how sensitive each MAUI is to variations in a particular health domain.

The ordinary least squares (OLS) estimator was used to estimate the models. In recent years, various econometric methods have been integrated into MAUI studies. These include the Tobit estimator and the censored least absolute deviations (CLAD) estimator. These techniques offer advantages, particularly in dealing with sample censoring of the dependent variable, such as many respondents reporting a utility value of 1. However, in the current analysis, these methods were not employed as sample censoring is not a significant concern. The descriptive statistics show that ceiling effects are minimal to moderate across instruments: 1.3% of SF-6D observations and 0.3% of AQoL-8D observations are at the maximum value, while the EQ-5D shows the highest proportion at 18%. The remaining instruments (HUI3, 15D, AQoL-4D) exhibit ceiling effects between 6.1 and 7.0%. All instruments demonstrate substantial variation across their utility ranges. Under these conditions, where censoring is limited and sample coverage is adequate, OLS produces unbiased estimates of mean predicted utility while providing superior coefficient interpretability and facilitating direct comparison with the existing mapping literature.

Furthermore, to evaluate how MAUIs respond to real-world health changes, we developed a series of simulated intervention scenarios based on common clinical treatments. These scenarios were informed by findings from previous studies and reflect typical changes in SF-36 dimensions associated with each intervention [2, 27, 30–32]. The specific change values for each SF-36 dimension were extracted from the cited studies, using reported mean differences between pre- and post-intervention scores. By selecting a diverse range of interventions, we sought to capture various aspects of HRQoL that are commonly affected in clinical practice. The following scenarios were selected for their clinical relevance, real-world applicability, and potential to reveal the interconnected nature of health outcomes:

1. Pain Relief Scenario: This scenario models an intervention focused on reducing bodily pain, leading to moderate improvements in physical functioning and vitality, alongside minor changes in mental health and other domains. Pain relief is essential for patients managing chronic pain, as it can enable greater participation in daily activities and enhance overall well-being. This scenario highlights the importance of pain management in improving both physical and mental aspects of HRQoL [27].
2. Depression Therapy Scenario: Reflecting psychological treatments such as cognitive behavioral therapy (CBT), this scenario captures the substantial improvements in mental health that can influence other health domains, including physical functioning, general health, vitality, and social functioning. This interconnected improvement emphasizes the bidirectional relationship between mental and physical health, where enhancements in mental well-being facilitate better physical and social functioning [30].
3. Cardiac Surgery Scenario: This scenario simulates the post-operative recovery process following cardiac surgery, characterized by short-term declines in role limitations and bodily pain due to recovery challenges, but with subsequent gains in vitality and slight improvements in physical functioning as recovery progresses. By illustrating both the immediate and long-term impacts of surgical interventions, this scenario provides insights into the recovery trajectories that affect HRQoL [32].
4. Stress Management Scenario: Stress management interventions, such as mindfulness or relaxation training, are captured in this scenario, which focuses on improvements in general health, mental health, vitality, and social functioning, with reductions in stress-related bodily pain. This scenario underscores the significant effects of reducing emotional strain on both mental and physical well-being, reflecting the widespread benefits of managing stress in daily life [2].
5. Exercise Scenario: Representing the benefits of regular physical activity, this scenario demonstrates considerable improvements in physical functioning, general health, vitality, and social functioning, with moderate gains in mental health and emotional role functioning. Exercise is known to enhance multiple health dimensions, providing a comprehensive boost to HRQoL [31].

These scenarios were chosen not only for their relevance to typical patient experiences but also for their ability to capture diverse health dimensions. Each scenario impacts multiple SF-36 domains, allowing us to assess how sensitively each MAUI reflects real-world health status changes across physical, mental, and social aspects of health. This comprehensive approach helps identify which MAUIs are best suited for capturing HRQoL changes, ultimately guiding their application in clinical and research settings. A summary of these scenarios and their associated changes in SF-36 dimensions is provided in Table 3.

To translate these changes into predicted differences in utility values, we applied the estimated regression coefficients (β) from the OLS models linking SF-36 dimensions to each MAUI in equation 1. For each simulated intervention, the expected change in utility (ΔUtility) was computed as the sum of the products of the observed mean change in each SF-36 dimension $\Delta SF36_i$ and its corresponding β coefficient. That is,

$$\Delta Utility_{MAUI} = \sum_{i=1}^{8} {}_i \times \Delta SF36_i$$

This approach allowed us to model how empirically observed changes in health status, expressed as differences in SF-36 scores, would translate into changes in utility values for each MAUI. In this way, the simulation provides an indicative assessment of instrument responsiveness and potential utility gains that could, in principle, inform incremental QALY estimation in cost-utility analyses.

To evaluate the performance of the models in this study, several statistical metrics were used. These metrics include root mean square error (RMSE), R-squared (R^2), and mean absolute error (MAE). RMSE provides a measure of the average magnitude of the errors between predicted and observed values, giving higher weight to larger errors. R^2, or the coefficient of determination, indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, thus assessing the model's explanatory power. MAE, on the other hand, measures the average absolute differences between predicted and actual values, offering a straightforward interpretation of prediction accuracy [33]. Together, these metrics offer a comprehensive evaluation of model performance, highlighting different aspects of predictive accuracy and reliability. All statistical analyses were performed in R version 4.3.2 [34].

Table 4 Summary statistics for multi-attribute utility instruments (MAUIs)

MAUI	Mean	SD	Min	Max
EQ-5D	0.728	0.231	−0.510	1.000
HUI3	0.698	0.274	−0.343	1.000
SF-6D	0.705	0.138	0.301	1.000
15D	0.844	0.130	0.253	1.000
AQoL-4D	0.626	0.268	−0.040	1.000
AQoL-8D	0.670	0.227	0.105	1.000

3 Results

The summary statistics for the MAUIs are reported in Table 4. The mean values vary across different MAUIs, ranging from 0.626 (AQoL-4D) to 0.844 (15D). The variability of these measures, indicated by their standard deviations, also differs, with HUI3 showing the greatest variability and SF-6D the least. All the MAUIs have a maximum value of 1.

Table 5 reports the Pearson correlation coefficients between MAUIs and SF-36 dimensions. SF-6D correlates strongly with dimensions such as Social Functioning (0.805) and Vitality (0.757), while AQoL-8D is particularly sensitive to Mental Health (0.784) and Vitality (0.798). Conversely, EQ-5D and HUI3 exhibit stronger associations with physical health dimensions like Bodily Pain (0.752 for EQ-5D and 0.670 for HUI3) and Physical Functioning, with comparatively lower sensitivity to emotional and role-related dimensions, such as Role Emotional. The 15D captures general health and vitality well but is less sensitive to emotional dimensions, while AQoL-4D shows moderate associations across dimensions, though it is somewhat limited in role emotional and role physical domains. Overall, SF-6D exhibits strong correlations with a majority of SF-36 dimensions, indicating its broad applicability for assessing HRQoL.

Table 3 Scenarios and corresponding changes in SF-36 dimension scores

Scenario	PF	RP	BP	GH	VT	SF	RE	MH
Pain relief	5.5	0.9	8.4	−0.8	3.1	6.1	0.6	1.6
Depression therapy	27.8	7.1	8.9	18.0	14.5	7.8	4.8	21.4
Cardiac surgery	2.2	−14.2	−7.5	1.2	18.1	−3.8	−9.5	0.7
Stress management	7.9	22.9	13.3	10.9	7.3	8.1	10.4	6.2
Exercise	14.0	10.0	8.0	12.0	9.0	13.0	6.0	7.0

SF-36 dimensions: *PF* Physical Functioning, *RP* Role Physical, *BP* Bodily Pain, *GH* General Health, *VT* Vitality, *SF* Social Functioning, *RE* Role Emotional, *MH* Mental Health

3.1 Unique Contributions of Individual SF-36 Dimensions to Observed MAUIs

Table 6 and Fig. 1 present the beta coefficients and R^2 from the regression of MAUI scores on single dimensions of the SF-36 from equation 1. In this table, beta coefficients represent the change in MAUIs associated with a one-unit change in the respective SF-36 dimension. R^2 values indicate the proportion of variance in MAUI explained by the SF-36 dimensions. Higher beta coefficients and R^2 values indicate stronger associations and better explanatory power, respectively.

The regression analyses highlighted distinct variations in the sensitivity of the six MAUIs to individual SF-36 dimensions, as shown by the beta coefficients and R^2 values in Table 6. 15D and SF-6D demonstrated strong sensitivities across multiple dimensions, particularly in RE, RP, and VT. In contrast, EQ-5D showed its strongest associations with PF (beta = 0.49, R^2 = 0.32) and BP (beta = 0.57, R^2 = 0.31), while having lower associations with dimensions related to emotional and mental health, such as RE (beta = 0.22) and MH (beta = 0.31), where R^2 values ranged from 0.30 to 0.33, indicating limited sensitivity in these areas. HUI3 exhibited a more balanced, moderate sensitivity across physical and mental health dimensions, with its strongest associations in PF (beta = 0.23, R^2 = 0.45), BP (beta = 0.26, R^2 = 0.45), and VT (beta = 0.34, R^2 = 0.40). Lower beta values for RP and SF suggest that HUI3 may be less sensitive to social or role-related limitations. AQoL-4D and AQoL-8D generally display lower sensitivity across SF-36 dimensions, with particularly low beta and R^2 values for MH and VT, while RE and RP show slightly higher beta values.

3.2 Most Influential SF-36 Dimensions in Determining HRQoL

Analysis from equation 2 identified MH, BP, and PF as the most influential dimensions across various quality-of-life indices represented by MAUIs (see Table 7 and Fig. 2). Notably, MH emerged as the most consistent factor, particularly strong in the HUI3, AQoL-4D, and AQoL-8D instruments.

In the EQ-5D model, PF, BP, GH, SF, and MH have significant positive associations with utility scores, with PF and BP contributing most strongly (β = 0.243 and 0.355, respectively). This model explains 71.0% of the variance (R^2 = 0.710), indicating a good fit, with RMSE and MAE

Table 5 Pearson correlation coefficients between multi-attribute utility instruments and SF-36 dimensions

Dimension	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
GH	0.633	0.631	0.684	0.725	0.636	0.699
PF	0.698	0.670	0.681	0.718	0.609	0.570
RP	0.555	0.539	0.708	0.610	0.535	0.541
BP	0.752	0.670	0.738	0.696	0.627	0.631
VT	0.617	0.632	0.757	0.722	0.660	0.798
SF	0.641	0.648	0.805	0.681	0.669	0.715
RE	0.468	0.477	0.697	0.529	0.500	0.590
MH	0.555	0.598	0.711	0.610	0.629	0.784

SF-36 dimensions: *PF* Physical Functioning, *RP* Role Physical, *BP* Bodily Pain, *GH* General Health, *VT* Vitality, *SF* Social Functioning, *RE* Role Emotional, *MH* Mental Health

Table 6 Beta coefficients and R^2 values from linear regression of multi-attribute utility instrument (MAUI) scores on SF-36 dimensions

	EQ-5D		HUI3		SF-6D		15D		AQoL-4D		AQoL-8D	
	Beta	R^2	Beta	R^2	Beta	R^2	Beta	R^2	Beta	R^2	Beta	R^2
GH	0.40	0.40	0.31	0.40	0.49	0.47	0.63	0.53	0.24	0.40	0.31	0.49
PF	0.49	0.32	0.23	0.45	0.47	0.46	0.61	0.52	0.21	0.37	0.34	0.32
RP	0.31	0.59	0.29	0.53	0.60	0.50	0.76	0.37	0.47	0.29	0.53	0.29
BP	0.57	0.31	0.26	0.45	0.46	0.55	0.63	0.48	0.22	0.39	0.33	0.40
VT	0.38	0.43	0.34	0.40	0.49	0.57	0.65	0.52	0.26	0.44	0.29	0.64
SF	0.41	0.31	0.20	0.42	0.39	0.65	0.60	0.46	0.12	0.45	0.21	0.51
RE	0.22	0.59	0.23	0.53	0.58	0.49	0.75	0.28	0.45	0.25	0.49	0.35
MH	0.31	0.33	0.19	0.36	0.40	0.51	0.60	0.37	0.10	0.40	0.12	0.61

SF-36 dimensions: *PF* Physical Functioning, *RP* Role Physical, *BP* Bodily Pain, *GH* General Health, *VT* Vitality, *SF* Social Functioning, *RE* Role Emotional, *MH* Mental Health

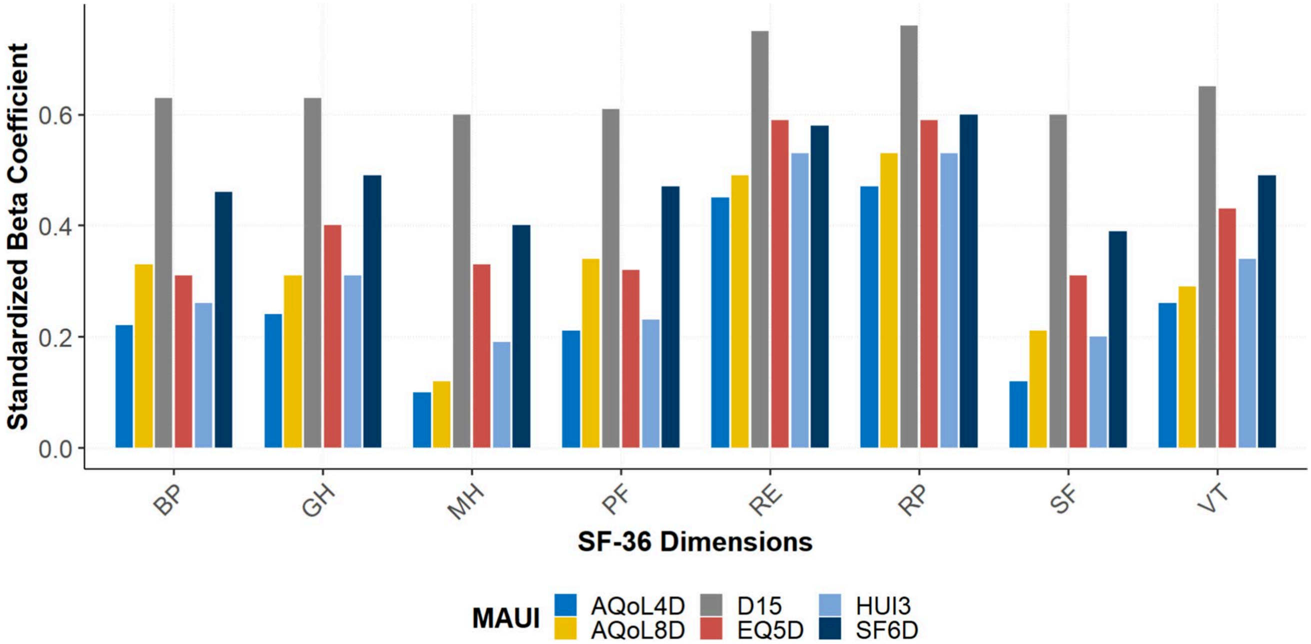

Fig. 1 Sensitivity of multi-attribute utility instruments (MAUIs) to SF-36 Dimensions (standardized beta coefficients)

values of 0.124 and 0.090, respectively. The HUI3 model shows statistical significance for all dimensions except RE, with PF, BP, GH, and MH as the strongest predictors (β = 0.305, 0.251, 0.093, and 0.342, respectively). This model accounts for 66.5% of the variance (R^2 = 0.665), with slightly higher RMSE and MAE values (0.159 and 0.117) compared with EQ-5D, indicating reduced predictive accuracy. The SF-6D model demonstrates high explanatory power (R^2 = 0.885), with all dimensions except GH contributing significantly. PF and BP have smaller coefficients relative to other MAUIs, while VT, SF, and MH show higher contributions. This model has the lowest RMSE (0.047) and MAE (0.037) among all instruments, highlighting its predictive precision. In the 15D model, BP (β = 0.135) and PF (β = 0.084) remain strong predictors, alongside significant contributions from GH, VT, and MH. This model explains 76.0% of the variance (R^2 = 0.760), with moderate error measures (RMSE = 0.064, MAE = 0.047), reflecting good performance. The AQoL-4D model identifies BP (β = 0.180) and SF (β = 0.157) as the most influential predictors, with

Table 7 Multivariate regression model with dimensions only

	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
Constant	0.089*	−0.062*	0.327*	0.494*	−0.086*	0.023*
PF	0.243*	0.305*	0.069*	0.135*	0.191*	0.075*
RP	−0.032*	−0.019*	0.041*	0.003	0.003	−0.009*
BP	0.355*	0.251*	0.119*	0.084*	0.18*	0.139*
GH	0.059*	0.093*	0.005	0.094*	0.127*	0.129*
VT	0.015	0.035*	0.073*	0.088*	0.093*	0.209*
SF	0.074*	0.107*	0.104*	0.042*	0.157*	0.057*
RE	0.01*	0.001	0.053*	0.013*	0.003	0.028*
MH	0.219*	0.342*	0.116*	0.088*	0.315*	0.384*
R-squared	0.710	0.665	0.885	0.760	0.642	0.808
RMSE	0.124	0.159	0.047	0.064	0.160	0.100
MAE	0.090	0.117	0.037	0.047	0.121	0.076

*Denotes statistical significance at 0.05 level

BP Bodily Pain, *GH* General Health, *MAE* mean absolute error, *MH* Mental Health, *PF* Physical Functioning, *RE* Role Emotional, *RMSE* root mean square error, *RP* Role Physical, *SF* Social Functioning, *VT* Vitality

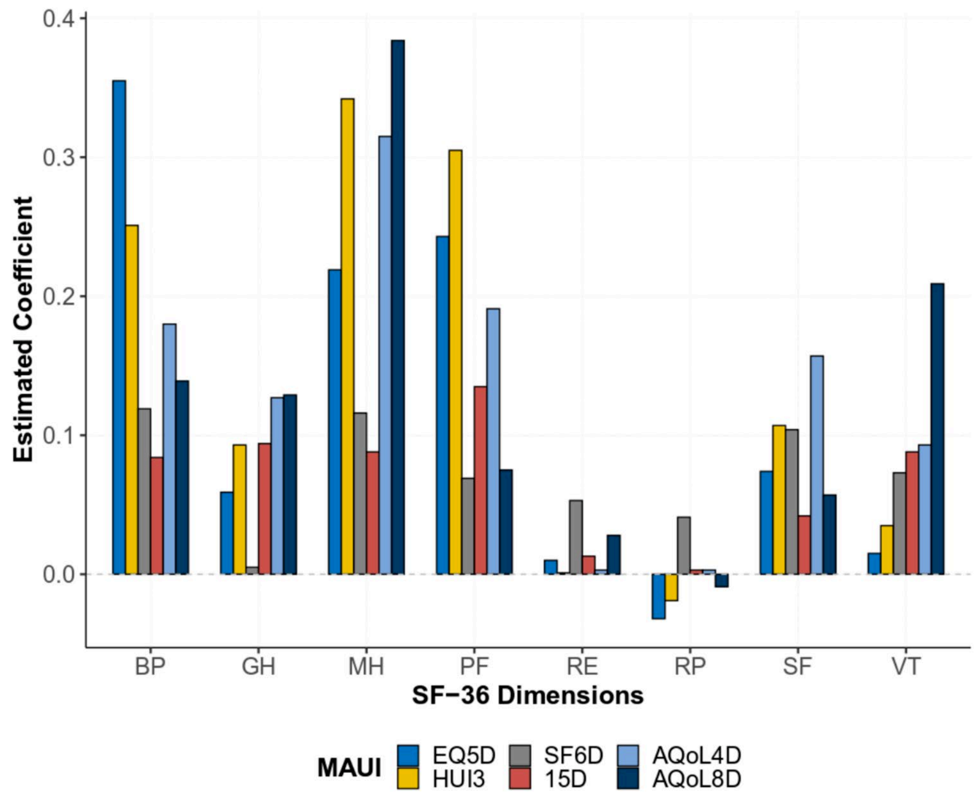

Fig. 2 Influence of SF-36 dimensions on multi-attribute utility instruments (MAUIs)

notable positive associations from PF and GH. This model has the lowest overall fit ($R^2 = 0.642$), with relatively high prediction errors (RMSE = 0.160, MAE = 0.121), indicating less effectiveness in explaining AQoL-4D utility scores. Finally, the AQoL-8D model reveals that MH ($\beta = 0.384$), BP ($\beta = 0.139$), and VT ($\beta = 0.209$) are the largest positive contributors to the utility score. With $R^2 = 0.808$, this model exhibits high explanatory power, and its prediction errors (RMSE = 0.100, MAE = 0.076) are lower than those of the AQoL-4D model, indicating better predictive accuracy. Table 8 summarizes the top three predictors for each MAUI based on the beta coefficients.

3.3 Instrument Sensitivity

To evaluate the sensitivity of each MAUI to variation in specific SF-36 dimensions, we examined the magnitude and direction of coefficients from the multiple linear regression models (Fig. 3). Each regression model estimates the utility score for a given MAUI as a function of the eight SF-36 domains. Sensitivity was defined as the magnitude of the standardized regression coefficient (β) for each SF-36 dimension. Larger absolute values of (β) indicate that the MAUI is more sensitive to change in that domain. Our analysis revealed varying sensitivities of the MAUIs to different dimensions of the SF-36.

Table 8 Top three influential predictors associated with each MAUI

MAUI	Top predictor		
	1	2	3
EQ-5D	BP (beta = 0.355)	PF (beta = 0.243)	MH (beta = 0.219)
HUI3	MH (beta = 0.342)	PF (beta = 0.305)	BP (beta = 0.251)
SF-6D	SF (beta = 0.104)	MH (beta = 0.116)	VT (beta = 0.073)
15D	PF (beta = 0.135)	GH (beta = 0.094)	VT (beta = 0.088)
AQoL-4D	MH (beta = 0.315)	SF (beta = 0.157)	BP (beta = 0.18)
AQoL-8D	MH (beta = 0.384)	VT (beta = 0.209)	BP (beta = 0.139)

BP Bodily Pain, *GH* General Health, *MAUI* multi-attribute utility instrument, *MH* Mental Health, *PF* Physical Functioning, *SF* Social Functioning, *VT* Vitality

For bodily pain, the EQ-5D exhibited markedly superior sensitivity, significantly outperforming the HUI3 and AQoL-4D, indicating it more effectively reflects changes in pain status. Regarding physical functioning, the HUI3 was the most sensitive instrument, followed by the EQ-5D and AQoL-4D. Conversely, the SF-6D and AQoL-8D showed comparatively lower sensitivities, potentially underrepresenting mobility-related health status. The AQoL-8D and AQoL-4D consistently ranked highest for general health sensitivity, followed by the HUI3. The SF-6D was the least sensitive to general health perceptions. Finally, all instruments exhibited minimal sensitivity to role physical, with coefficients approaching zero. The EQ-5D and HUI3 even showed small negative coefficients, indicating a weak or potentially counterintuitive association within this dataset.

For mental health, the AQoL-8D and HUI3 demonstrated the strongest sensitivity, followed by the AQoL-4D and EQ-5D. This suggests these instruments are particularly adept at capturing variations in psychological well-being, which aligns with their broader inclusion of mental health content. The AQoL-8D also stood out for its sensitivity to vitality, preceding the AQoL-4D and 15D. Notably, the EQ-5D was largely unresponsive to vitality, underscoring concerns about its limited health coverage in this domain. For role emotional functioning, the SF-6D displayed the greatest sensitivity, with the AQoL-8D and 15D showing smaller but significant sensitivities. The EQ-5D, HUI3, and AQoL-4D demonstrated minimal or near-zero coefficients, suggesting limited ability to capture changes in this dimension. Concerning social functioning, the AQoL-4D and SF-6D were the most sensitive, while the EQ-5D and AQoL-8D showed relatively lower coefficients.

In summary, each MAUI displayed a distinct profile of sensitivity across the SF-36 dimensions. These findings underscore that no single MAUI is uniformly sensitive across all SF-36 domains. Therefore, the selection of an instrument should be carefully considered and align with the specific health impacts and domains of interest in a given study or intervention.

3.4 Scenario Analysis

Table 9 and Fig. 4 present the estimated impact of the different intervention scenarios, outlined in Sect. 2.2, on quality-of-life scores across MAUIs. The values shown represent the predicted average change in utility scores when patients experience the specific health improvements associated with each intervention. These scenarios were designed to reflect realistic clinical outcomes based on published research, including treatments for pain relief, depression therapy, cardiac surgery recovery, stress management programs,

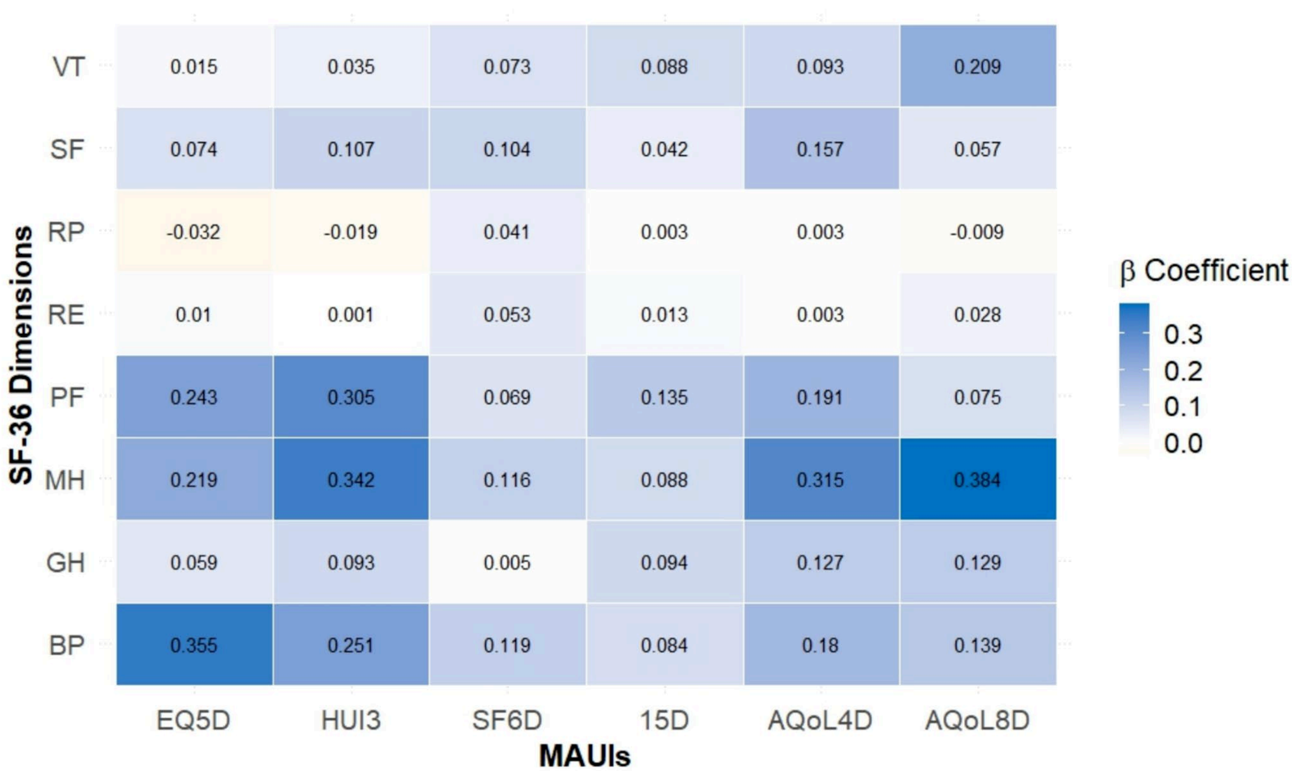

Fig. 3 MAUI sensitivity to SF-36 dimensions (multivariate regression coefficients). *MAUI* multi-attribute utility instrument

and exercise interventions. By applying the statistical relationships we established between SF-36 health dimensions and each utility instrument, we can estimate how much each instrument would detect improvements (or temporary declines) in patients' quality of life following these interventions.

Values represent predicted changes in utility scores expressed as percentages (0–100 scale). For example, a value of 5.111 represents a utility change of 0.051 on the standard 0–1 utility scale

Depression Therapy yielded the most significant improvements across all instruments, with the highest sensitivity observed in HUI3 (+20.928) and AQoL-8D (+17.435). EQ-5D (+16.279) and SF-6D (+7.936) also showed substantial gains, though smaller in magnitude. Pain Relief demonstrated moderate improvements across all instruments, with the largest changes seen in EQ-5D (+5.111) and HUI3 (+5.034). Stress Management and Exercise both resulted in notable positive changes, particularly in HUI3 (+9.590, +11.316) and AQoL instruments. In contrast, Cardiac Surgery showed a minimal or slightly negative impact on quality of life, with the largest decline observed in EQ-5D (+0.756) and HUI3 (+1.432). However, AQoL-8D (+4.237) and 15D (+1.968) captured modest positive changes. Overall, the results indicate that HUI3, AQoL-4D, and AQoL-8D are the most sensitive instruments to mental health interventions, while EQ-5D and HUI3 are more sensitive to physical interventions like pain management and exercise. Cardiac Surgery had the smallest overall impact across instruments, with minimal improvements in quality of life in several MAUIs and a notable benefit in AQOL-8Ds.

Table 9 Predicted mean changes in multi-attribute utility instruments (MAUIs) for different scenarios

Scenario	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
Pain relief	5.111	5.034	2.488	2.072	4.252	3.132
Depression therapy	16.279	20.928	7.936	9.743	18.541	17.435
Cardiac surgery	0.756	1.432	1.353	1.968	2.246	4.237
Stress management	8.732	9.590	5.746	4.928	9.289	8.317
Exercise	9.321	11.316	5.512	5.738	10.767	9.111

Values represent predicted changes in utility scores expressed as percentages (0-100 scale). For example, a value of 5.111 represents a utility change of 0.051 on the standard 0-1 utility scale.

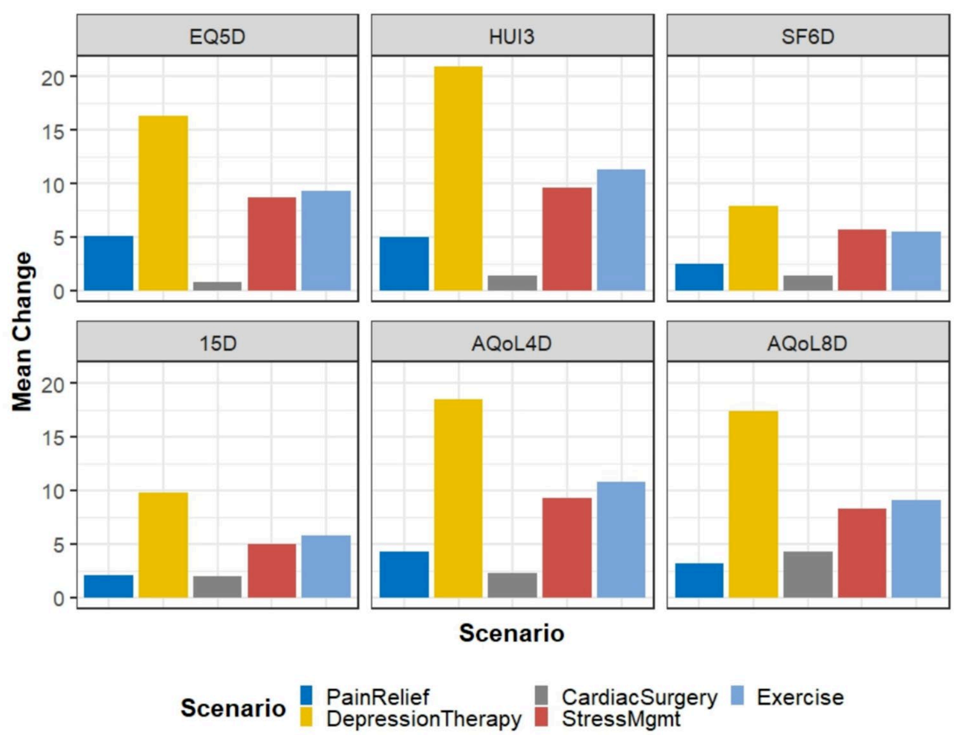

Fig. 4 Predicted mean changes across MAUIs for different scenarios. Y-axis values are on a 0–100 scale. To convert to standard utility scale (0–1), divide by 100

4 Discussion

This study reveals significant variations in the sensitivity of six widely used multi-attribute utility instruments (MAUIs)—EQ-5D, SF-6D, HUI3, 15D, AQoL-4D, and AQoL-8D—to the eight SF-36 health dimensions, highlighting the importance of aligning instrument choice with the specific health domains most relevant to the intervention being studied. Each MAUI exhibited unique patterns of sensitivity, suggesting that the selection of an instrument can substantially influence the measurement of HRQoL outcomes according to the target population and intervention type.

Our findings build on and extend the work of Richardson and colleagues (MIC studies), who examined the comparative performance of MAUIs in the same dataset. Consistent with their results, we found notable variation in sensitivity across instruments; however, our study provides additional insight by focusing on dimension-level associations rather than overall summary correlations. Furthermore, the inclusion of simulated interventions demonstrates how differential sensitivity across MAUIs may translate into varying QALY estimates, thereby linking psychometric sensitivity to practical implications for health economic evaluation.

The results of our analysis demonstrate that different SF-36 dimensions contribute in varying degrees to each MAUI's utility scores. As shown in Table 5, the SF-6D behaves as a distinct preference-based measure rather than a simple rescaling of the SF-36. While it retains meaningful associations with SF-36 dimensions, the valuation-based transformation produces a conceptually different construct which is consistent with the purpose of preference-based scoring. Consequently, correlations between the SF-6D and SF-36 subscales, though moderate to strong (e.g., VT: 0.757; SF: 0.805), are not substantially higher than those for other MAUIs, reflecting the SF-6D's reduced item set and compressed domain structure. As the SF-6D captures only a portion of the original SF-36 content, perfect or near-perfect correlations should not be expected. Furthermore, 15D and SF-6D showed broad sensitivity across multiple SF-36 domains, particularly in social functioning, role limitations, vitality, and mental health. This finding suggests that both the 15D and SF-6D provide a more comprehensive assessment of HRQoL, particularly when studies focus on both physical and mental health domains, including social and role-oriented outcomes [11, 27]. In contrast, EQ-5D showed a stronger association with physical health dimensions, particularly bodily pain and physical functioning, but lower associations with mental and role-related domains. This pattern is consistent with earlier findings that EQ-5D may be less suited to capture the complex aspects of mental and social health, thereby potentially limiting its applicability in studies focused on psychological well-being or emotional health [35, 36].

HUI3 demonstrated balanced sensitivity across physical and mental health dimensions, with strong associations in PF, BP, and VT, which supports its versatility in HRQoL measurement across a variety of health conditions. However, its lower sensitivity to role limitations and social functioning suggests that while HUI3 may be useful in physical health evaluations, it may have limitations in capturing the full range of social and role-based quality of life aspects compared with SF-6D and 15D [37, 38]. AQoL-4D and AQoL-8D, although generally showing lower overall sensitivity, demonstrated notable sensitivity to mental health, supporting their suitability in contexts where mental health is a central concern, such as interventions targeting psychological or social well-being [14, 23].

Regression models further revealed the SF-36 mental health (MH), bodily pain (BP), and physical functioning (PF) dimensions as top predictors of utility scores across most MAUIs, with MH emerging as the most influential predictor overall. The prominence of MH across instruments, particularly AQoL-8D, highlights the need for robust mental health measurement in HRQoL assessments. Similarly, PF showed particular relevance for EQ-5D and 15D, while BP was influential for both EQ-5D and HUI3. The relatively lower contributions of role limitations and social functioning imply that these dimensions, while important, may not be central to the utility measurements of these instruments. Thus, interventions that aim to enhance physical functioning, pain management, and mental health are likely to achieve the most substantial improvements in utility scores and overall quality of life.

Simulated health intervention analyses further clarified each instrument's sensitivity to specific types of health improvements. Mental health interventions, such as depression therapy, were best captured by HUI3, AQoL-4D, and AQoL-8D, which demonstrated strong sensitivities to improvements in mental and social health domains. This suggests that these instruments may be particularly effective for evaluating psychological and social outcomes and, thus, highly suitable for mental health interventions. Conversely, pain relief interventions yielded the most significant changes in EQ-5D and HUI3 scores, indicating these instruments' strengths in measuring physical functioning and pain-related domains [8, 12]. Stress management and exercise interventions, which involve both mental and physical improvements, showed high sensitivity in AQoL instruments and HUI3, reinforcing the versatility of these instruments for capturing a broad range of HRQoL dimensions [19]. The complex response patterns observed following cardiac surgery recovery suggest that EQ-5D and HUI3 are highly sensitive to immediate physical challenges post-surgery, while

15D and AQoL-8D may be better suited to capture the more gradual recovery process in HRQoL that includes physical, social, and mental domains.

While our analysis focused on cross-sectional sensitivity, the simulation framework also provides insight into instrument responsiveness—the extent to which utility scores change in response to meaningful health improvements. By applying regression-derived coefficients to empirically observed SF-36 changes, we estimated the expected utility gains for typical clinical interventions. This approach conceptually parallels the generation of incremental QALYs in cost-utility analyses, since predicted ΔUtility values represent potential changes in HRQoL over time. Although not a longitudinal design, these simulations illustrate how instrument choice could influence estimated QALY gains in HTA contexts, given differing responsiveness profiles across MAUIs. Future studies incorporating longitudinal data could further validate these responsiveness estimates and their translation into incremental QALY outcomes.

Beyond measurement considerations, these findings have important implications for healthcare decision making and cost-effectiveness evaluation. In cost-utility analyses, the choice of MAUI directly affects utility scores and, consequently, the incremental quality-adjusted life years (QALYs) derived from interventions. Because MAUIs differ in their domain sensitivities, the same clinical improvement can yield varying utility gains depending on the instrument used. For instance, instruments that are more sensitive to physical health domains (e.g., EQ-5D, HUI3) tend to produce larger utility gains for interventions targeting mobility or pain, whereas those more responsive to mental and social dimensions (e.g., AQoL-8D, SF-6D) may capture greater benefits in psychological or community-based programs. As a result, two interventions with comparable clinical outcomes could produce different incremental cost-effectiveness ratios (ICERs), potentially influencing reimbursement and policy decisions.

It is also important to recognize that although all MAUIs are anchored at 0 (dead) and 1 (full health), they differ in scale distribution and interval properties, as highlighted by Richardson et al. [12]. These scale characteristics can influence both the apparent sensitivity of instruments to SF-36 dimensions and the magnitude of derived utility changes. For example, instruments with compressed scales in poorer health states (such as the SF-6D) may appear less sensitive to severe health decrements than those with broader scaling (e.g., EQ-5D). Accordingly, cross-instrument comparisons of sensitivity and QALY outcomes should be interpreted with caution, acknowledging that part of the observed variation reflects inherent scaling differences.

In addition, health state valuation systems undergo constant refinement and adjustment as methodological standards and population preferences evolve. Consequently, one consideration in interpreting these findings is the use of the EQ-5D-5L crosswalk [20] and the original SF-6D valuation [27], rather than newer direct value sets or the SF-6Dv2. This choice was made to maintain internal consistency with the Multi-Instrument Comparison (MIC) study database [12, 19], which served as the data source for this research. The MIC study utilized these specific frameworks to enable standardized comparisons across multiple international cohorts. While more recent valuation methods are now available, employing the original MIC study value sets ensures our results remain directly comparable with the extensive body of psychometric evidence derived from this dataset.

Previous studies further illustrate the practical implications of these instrument differences. Sach et al. [39] found that using SF-6D instead of EQ-5D for knee replacement increased mean utility values but reduced estimated QALY gains, resulting in higher ICERs. Similarly, Campbell et al. [40] reported that AQoL instruments detected larger utility changes than EQ-5D in mental health contexts, potentially making such interventions appear more cost effective. These examples underscore the need to align MAUI selection with the dominant health domains affected by an intervention to ensure accurate and equitable resource allocation.

Taken together, our comparative results provide an empirical basis for guiding instrument choice in HTA and cost-utility analysis. By linking domain-level sensitivity to potential QALY impacts, this study supports more informed and transparent decision making in economic evaluation and policy contexts. Overall, these findings highlight the importance of carefully selecting MAUIs based on the targeted health domains of interest and the specific characteristics of the population under study. Choosing the most appropriate instrument based on its sensitivity to relevant health dimensions can significantly enhance the accuracy and validity of HRQoL measurement, leading to more reliable insights into patient outcomes. Future research should continue to explore these sensitivity patterns across diverse populations and clinical contexts, thereby refining the selection criteria for MAUIs and optimizing their application in HRQoL research and clinical practice.

5 Limitations

This study has several limitations that should be acknowledged. First, while our analysis focused on cross-sectional sensitivity patterns, the simulation framework provides only indirect insight into instrument responsiveness. Future studies incorporating longitudinal data could further validate these responsiveness estimates and their translation into incremental QALY outcomes. Second, the data used in this study were collected in the early 2010s, and while the fundamental properties of these instruments remain relevant, changes in population health characteristics or valuation preferences over time may affect the generalizability of our findings to current contexts. Third, our simulated interventions were based on observed SF-36 changes from the literature and regression-derived coefficients rather than actual intervention trials within the MIC dataset. Although this approach provides valuable insights into potential instrument behavior, direct longitudinal assessment would offer more definitive evidence of responsiveness. Finally, our analysis was limited to six MAUIs and the SF-36 framework; future research should continue to explore these sensitivity patterns across diverse populations, clinical contexts, and other HRQoL measurement frameworks, so as to refine the selection criteria for MAUIs and optimizing their application in HRQoL research and clinical practice.

6 Conclusion

This study analyzes the sensitivity of six widely used multi-attribute utility instruments (MAUIs) to the SF-36 health dimensions, highlighting distinct patterns of sensitivities across instruments. Findings reveal that each MAUI varies in sensitivity to specific health domains, making careful instrument selection essential to accurately capture outcomes relevant to a given intervention. For example, in mental health contexts, the HUI3 and AQoL-8D are particularly sensitive due to their heightened sensitivity to mental and social functioning. In contrast, physical health interventions, such as pain management or rehabilitation programs, might benefit from the EQ-5D or HUI3, which are more attuned to changes in physical functioning and bodily pain.

The study also underscores that each MAUI's utility scores are influenced differently by the SF-36's eight dimensions, enabling researchers to better understand which dimensions most strongly drive the results. This nuanced insight supports a more informed approach to MAUI selection, aligning the choice of instrument with the health dimensions most pertinent to the study.

Future research should further investigate these sensitivity patterns across diverse populations and clinical contexts. By identifying the most sensitive instruments for different health domains, researchers can optimize MAUI use in HRQoL assessments, enhancing the precision and relevance of their findings.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Materials Data are available upon request.

Code Availability Not applicable.

Competing Interests The authors declare that there is no conflict of interest.

Funding No funding was received for this publication.

Authors' Contributions The research was conceptualized by MS. All authors contributed substantially to the study design. JT conducted the data analysis and interpretation and drafted the initial manuscript. MS critically revised the manuscript for important intellectual content. All authors reviewed, provided feedback on, and approved the final version submitted for publication.

References

1. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118(8):622–9.
2. Michalsen A, Grossman P, Lehmann N, Knoblauch NTM, Paul A, Moebus S, et al. Psychological and quality-of-life outcomes from a comprehensive stress reduction and lifestyle program in patients with coronary artery disease: results of a randomized trial. Psychother Psychosom. 2005;74(6):344–52.
3. Lau VI. Health-related quality-of-life and health-utility reporting in critical care. World J Crit Care Med. 2022;11(4):236–45.
4. Lau VI. Health-related quality-of-life and cost utility analyses in critical care: a systematic review. Crit Care Med. 2021;49(4):575–88.
5. Drummond MF. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.
6. Chen G, Ratcliffe J. A review of the development and application of generic multi-attribute utility instruments for paediatric populations. Pharmacoeconomics. 2015;33(10):1013–28.
7. Kennedy-Martin M. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. Eur J Health Econ. 2020;21(8):1245–57.
8. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. Appl Health Econ Health Policy. 2017;15(2):127–37.
9. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33(5):337–43.

10. Brooks R, Group E. EuroQol: the current state of play. Health Policy. 1996;37(1):53–72.
11. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med. 2001;33(5):328–36.
12. Richardson J, Khan MA, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI3, 15D, QWB, and AQoL-8D multiattribute utility instruments. Med Decis Making. 2015;35(3):276–91.
13. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res. 2008;17:485–95.
14. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Ann Med. 2001;33(5):358–70.
15. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. Ann Med. 2001;33(5):375–84.
16. Ware JE Jr. SF-36 health survey update. Spine (Phila Pa 1976). 2000;25(24):3130–9.
17. Ware JE Jr., Sherbourne CD, Mazel RM. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993;31(3):247–63.
18. Richardson J. Measuring the sensitivity and construct validity of 6 utility instruments in 7 disease areas. Med Decis Making. 2016;36(2):147–59.
19. Richardson J. Cross-national comparison of twelve quality of life instruments. MIC Paper. 2014;2:1–33.
20. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708–15.
21. Feeny D. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. Med Care. 2002;40(2):113–28.
22. Grutters JP. Choosing between measures: comparison of EQ-5D, HUI2 and HUI3 in persons with hearing complaints. Qual Life Res. 2007;16:1439–49.
23. Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual Life Res. 1999;8:209–24.
24. Brazier J. Measuring and valuing health benefits for economic evaluation. 2nd ed. Oxford: Oxford University Press; 2017.
25. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling the utility of health states with the Assessment of Quality of Life (AQoL)-8D instrument: overview and utility scoring algorithm. Research Paper 63, Centre for Health Economics; 2011.
26. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727–36.
27. Brazier J, 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.
28. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ. 1993;2(3):217–27.
29. Ware JE Jr. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute; 1993.
30. Rogers WH, Wittink H, Wagner A, Cynn D, Carr DB. Assessing individual outcomes during outpatient multidisciplinary chronic pain treatment by means of an augmented SF-36. Pain Med. 2000;1(1):44–54.
31. Kennedy SH, Eisfeld BS, Cooke RG. Quality of life: an important dimension in assessing the treatment of depression? J Psychiatry Neurosci. 2001;26(Suppl):S23.
32. Nolte K, Herrmann-Lingen C, Wachter R, Gelbrich G, Düngen HD, Duvinage A, et al. Effects of exercise training on different quality of life dimensions in heart failure with preserved ejection fraction: the Ex-DHF-P trial. Eur J Prev Cardiol. 2015;22(5):582–93.
33. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 2005;30(1):79–82.
34. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2021.
35. Brazier J. Is the EQ-5D fit for purpose in mental health? Br J Psychiatry. 2010;197(5):348–9.
36. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess. 2014;18(34).
37. Fitzpatrick R. Quality of life measures in health care. I: applications and issues in assessment. Br Med J. 1992;305(6861):1074–7.
38. Luo N, Feeny D, Coons SJ. A study of the construct validity of the Health Utilities Index Mark 3 (HUI3) in patients with schizophrenia. Qual Life Res. 2006;15:889–98.
39. Sach TH, Barton GR, Jenkinson C, Doherty M. Comparing cost-utility estimates: does the choice of EQ-5D or SF-6D matter? Med Care. 2009;47(8):889–94.
40. Campbell JA, Devlin NJ, Parkin D. A head-to-head comparison of the EQ-5D-5L and AQoL-8D multi-attribute utility instruments in patients who have previously undergone bariatric surgery. Patient. 2016;9(4):311–22.