Comparing the measurement properties of the EQ-5D-5L, SF-6Dv2, QLU-C10D and FACT-8D among survivors of Hodgkin's lymphoma

Richard Xu[a], Zoey Zhao[a], Tianxin Pan[b], Dong Dong[c]

a. *Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hong Kong, China*
b. *Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia*
c. *JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China*

Abstract

Objective: This study aimed to evaluate the psychometric properties of EQ-5D-5L, SF-6Dv2, QLU-C10D, and FACT-8D in survivors of Hodgkin's Lymphoma (HL). **Methods:** A cross-sectional, web-based survey was conducted from May to August 2022 to collect data. Chinese value sets were used to estimate the utility scores for EQ-5D-5L, SF-6Dv2, and QLU-C10D, while the Australian value set was used for FACT-8D. The measurement properties assessed included ceiling and floor effects, convergent validity (assessing associations between similar dimensions/utility scores from the four measures using Spearman's rank correlation and intraclass correlation coefficient), and known-group validity (determining whether the measures could differentiate health-related quality of life (HRQoL) between risk groups). **Results:** A total of 534 HL survivors participated in the survey and completed the questionnaire. The mean scores (SD) for EQ-5D-5L, SF-6Dv2, QLU-C10D, and FACT-8D were 0.89 (0.16), 0.71 (0.19), 0.72 (0.24), and 0.58 (0.18), respectively. All dimensions of EQ-5D-5L, SF-6D (except for vitality), QLU-C10D, and FACT-8D showed ceiling effects, ranging from 18% to 91.6%. In terms of convergent validity, all 30 pairs of associations between similar dimensions from the four measures were statistically significant, with correlation coefficients ranging from 0.29 to 0.77. Regarding utility scores, the EQ-5D-5L utility score showed a stronger correlation with SF-6Dv2 than with the other two condition-specific patient-reported measures (PRMs). Significant correlations of utility scores between the four measures were observed, with the intraclass correlation coefficient (ICC) values ranging from 0.23 to 0.73. EQ-5D-5L can significantly

differentiate HRQoL among all risk groups, while SF-6Dv2, QLU-C10D, and FACT-8D showed a less strong discriminant ability. **Conclusions:** EQ-5D-5L outperformed SF-6Dv2 in terms of agreement with cancer-specific PRMs and discriminant ability. However, SF-6Dv2 showed stronger associations with similar dimensions of QLU-C10D and FACT-8D, indicating high convergent validity. The selection of PRMs will affect the estimation of QALYs to support economic evaluation.

Key words: EQ-5D; SF-6D; QLU-C10D; FACT-8D; preference-based measure; psychometric properties

Introduction

Hodgkin's lymphoma (HL) is a type of cancer that affects the lymphatic system. It is a leading cause of cancer-related deaths in adults and its incidence rates are increasing worldwide. Although HL can affect people of all ages, it is more common in those over 15 years of age. Approximately 10,900 and 1,200 new cases of HL are diagnosed annually in the US and UK, respectively. In China, more than 8,000 new cases of HL are reported each year. HL is prevalent among people under 50 years of age, accounting for approximately one-third of all cancers in this age group. Patients with HL typically experience painless swelling of the lymph nodes in their neck or chest regions. They may also present with weight loss and fever. Common symptoms of HL include night sweats, fatigue and unintentional weight loss.

In recent years, new treatments for HL have been introduced. For example, brentuximab vedotin has been shown to significantly improve remission in children and adolescents with HL. In addition, targeted immunotherapy has revolutionised the treatment of HL. For instance, chimeric antigen receptor T-cell therapy is being investigated in clinical trials as a potentially effective treatment option [1]. Economic evaluation plays a crucial role in assessing the cost-effectiveness and value of interventions for HL because health-care decision-makers must understand the economic implications of different treatment options and allocate their resources efficiently. Furthermore, HL is a rare type of lymphoma with unique histological, immunophenotypic and clinical features. Relying solely on traditional clinical measures may not fully capture the holistic impact of the disease on patients' lives. Health-related quality of life (HRQoL) is a multidimensional concept that captures the impact of a disease on patients' overall health and well-being. HRQoL reflects the variations in the health status of patients with lymphoma who have different clinical conditions and socioeconomic characteristics at various stages of disease care (with regard to their diagnosis, treatment and prognosis). In addition to commonly used patient-reported outcome measures, HRQoL can also be assessed using preference-based measures (PBMs). These may be used to estimate the longitudinal effects of a disease on HRQoL and to calculate the number of quality-adjusted life years for health economic evaluations to facilitate decision-making regarding resource allocation [2].

The EQ-5D-5L and SF-6D are the two most widely used generic PBMs. The EQ-5D-5L consists of five generic health domains, with each domain assessed using a single question answered by the patient using a 5-level response scale. This measure has been applied to various health conditions to support rational decision-making regarding resource allocation in the global health-care sector [3]. The original SF-6D (SF-6Dv1) was derived from the 36-item Short-Form Health Survey (SF-36) and includes six dimensions [4]. The most recent update, SF-6Dv2, has undergone significant revisions to address ambiguities in its dimension levels and to unify the inconsistently positive and negative wording in SF-6Dv1 [5]. It is important to note that the measurement properties of these measures vary across different health conditions, although they are widely used in different populations. In the context of cancer research, a key concern is the limited sensitivity of these measures to capture the relevant health issues due to the restricted number and type of dimensions [6].

Condition-specific measures are increasingly used to assess HRQoL in cancer clinical trials. One such measure is the Functional Assessment of Cancer Therapy – Eight Dimensions (FACT-8D) [7], which is derived from the cancer-specific HRQoL profile measure known as the Functional Assessment of Cancer Therapy – General (FACT-G) [8]. The FACT-G is widely used in oncology clinical trials, either as a stand-alone questionnaire or included within specific modules. Another cancer-specific measure is the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Utility Core – 10 Dimensions (QLU-C10D) [9], which focuses on the 10 dimensions derived from the EORTC QLQ-C30, one of the most commonly used instruments for assessing HRQoL in patients with cancer.

Utility scores vary among PBMs due to differences in their structure and valuation methods [10]. However, there is a lack of psychometric evidence to compare the performance of cancer-specific PBMs to generic PBMs in HL. Considering the widespread use of the QLQ-C30 and FACT-G in cancer studies, it is important to determine if the performance of the two newly developed PBMs is appropriate when compared to the well-established generic PBMs for economic evaluation. Evidence for the measurement properties of PBMs in HL can provide valuable information for the application of these instruments to support the increasing need for economic evaluation in this population. Therefore, the objective of this study was to evaluate the

psychometric properties of the EQ-5D-5L, SF-6Dv2, QLU-C10D and FACT-8D in survivors of HL.

Methods

Participants and data source

The data used for this analysis were obtained from a nationwide cross-sectional online survey that aimed to understand the HRQoL and social and health status of individuals with HL in China from September to November 2022. The survey was conducted with the assistance of House086, which is one of the largest organisations that serves individuals with lymphoma and their families in China. All respondents were House086 members. The inclusion criteria were as follows: 1) age ≥ 18 years, 2) able to read and write in Chinese, 3) no cognitive problems and 4) able to provide informed consent. We recruited the participants via House086's internal social media platform, which uses both WeChat and QQ. Registered members received the study information and expressed interest in participating in the survey. Eligible members were invited to join a specific online survey group. Subsequently, staff members from the research team and House086 informed the respondents about the study and administered the survey. All respondents provided their informed consent before they joined the survey. This study was approved by the Institutional Review Board of the Chinese University of Hong Kong (Ref. No.: SBRE-18–268).

Measures

EQ-5D-5L

The EQ-5D-5L comprises two sections, where the first section is a health state classification system with five dimensions, namely, mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has five response levels that range from 'no problem' to 'extreme problems'. All health states described by the classification system can be summarised as utility scores ranging from 0 (death) to 1 (full health), where a negative score indicates a health state worse than death. In this study, the EQ-5D-5L utility score was estimated using Chinese value set [11], which ranges from −0.391 to 1.0. The second section is the Visual Analogue Scale (EQ VAS), which ranges from 0 (worst imaginable health) to 100 (best imaginable health) and represents individuals' global assessment of health.

SF-6Dv2

The SF-6Dv2 is derived from 10 items in the SF-36. The health state classification system of SF-6Dv2 comprises six dimensions, namely, physical functioning, role limitation, social functioning, pain, mental health and vitality. The pain dimension has six response levels, while all others have five levels. The Chinese SF-6Dv2 value set was used in this study to estimate the utility score, which ranges from −0.277 to 1 [12].

EORTC QLU-C10D

Derived from the EORTC QLQ-C30, the QLU-C10D has 10 items that select and combine four functional domains (i.e. physical, role, social, emotional) and six symptoms (i.e. pain, fatigue, sleep, appetite, nausea, bowel problems) from the QLQ-C30. Each QLU-C10D item has four levels including 'not at all' (level 1), 'a little' (level 2), 'quite a bit' (level 3) and 'very much' (level 4). In our study, the C10D utilities were estimated using a Hong Kong Chinese value set [13], where the utility scores range from −0.169 to 1.

FACT-8D

The FACT-8D comprises eight dimensions, namely, pain, fatigue, nausea, sleep, work, support, sadness and worry, which are derived from nine FACT-G items. Each dimension in FACT-8D has the following five response options: not at all (0), a little bit (1), somewhat (2), quite a bit (3) and very much (4). The FACT-8D encompasses more than 390,000 possible health states (5^8 = 390,625). However, there is currently no available Chinese value set for FACT-8D. In this study, the Australian value set was used to estimate the utility score [14].

Data analysis

Descriptive analysis was used to describe the patients' background characteristics and health statuses. The EQ-5D, SF-6D, QLU-C10D and FACT-8D profiles, including the mean, standard deviation, ceiling effect (i.e. percentage of highest possible scores) and floor effect (i.e. percentage of lowest possible scores) were obtained. More than 15% of the sample reporting the lowest or highest score was indicative of a floor or ceiling effect, respectively.

Convergent validity was assessed using hypothesis testing. Thirty pairs of correlations between the four measures were hypothesised. For example, we expected a moderate-to-strong correlation

between mobility assessed using the EQ-5D and physical functioning assessed using the SF-6D, a moderate-to-strong correlation between anxiety/depression assessed using the EQ-5D and emotional functioning assessed using the QLU-C10D, and a moderate-to-strong correlation between nausea assessed using QLU-C10D and FACT-8D. Spearman's correlation coefficient (ρ) was used to assess the strength of the hypotheses (weak, $\rho \leq 0.3$; moderate, $0.31 \leq \rho < 0.5$; or strong, $\rho \geq 0.5$) [36]. The intra-class correlation coefficient (ICC) was used to assess the agreement between the utility scores from the four measures (ICC > 0.5, moderate; ICC > 0.75, good). In addition, the agreement between measures (EQ-5D vs. SF-6D, EQ-5D vs. QLU-C10D, EQ-5D vs. FACT-8D, SF-6D vs. QLU-C10D, SF-6D vs. FACT-8D and QLU-C10D vs. FACT-8D) was assessed based on the ICC (> 0.7, satisfactory) and Bland–Altman plots. A bootstrap method (resamples = 1,000) was used to calculate the robust 95% confidence interval of the coefficient. Known-group validity was examined using analyses of variance. We hypothesised that at-risk participants would be more likely to report a lower HRQoL (i.e. utility score). F-statistics were calculated to assess the discriminant efficiency of the measures in differentiating between participants with different conditions.

Results

Participants' background characteristics

Table 1 presents the background characteristics of all patients who participated in this survey, of whom 48.5% were men, 59.7% were married and 45.3% were actively employed. The most common diagnosed cancer stage was IIA (33.1%), followed by stage IV, indicating a significant number of late-stage cancer diagnoses. The patients' ages ranged from 18 to 82 years, with an average age of 35.6 years. Additionally, a majority of the participants (71.9%) held a tertiary degree or higher.

Ceiling and floor effects

Table 2 presents the distribution of responses for the four measures. Strong ceiling effects were observed for all five dimensions of the EQ-5D-5L, ranging from 40.8% (i.e. anxiety/depression) to 91.6% (i.e. self-care). For the SF-6Dv2, ceiling effects were observed for five out of six dimensions, except for vitality (9.4%). In addition, four dimensions (i.e. physical functioning, role limitation, social functioning, and vitality) out of six showed slight-to-moderate floor

effects, ranging from 19% to 29%. For the two cancer-specific measures, ceiling effects were observed for all dimensions of the QLU-C10D and FACT-8D. For the QLU-C10D, participants most often selected 'no problem' for role limitation (65.4%), followed by 'appetite' (62.4%) and 'nausea' (65.2%). The strength of the ceiling effect was lower for the FACT-8D than the QLU-C10D, with slightly more than half of the respondents reporting no problem with 'support' and 'pain' in the FACT-8D. Approximately 18% of participants reported no problem with 'sleep', which was the lowest percentage for all dimensions from the FACT-8D. Furthermore, the mean (standard deviation) value of the EQ-5D-5L utility and EQ-VAS scores were 0.89 (0.16) and 78.4 (17.4), respectively. Likewise, the utility scores for the SF-6Dv2, QLU-C10D and FACT-8D were 0.71 (0.19), 0.72 (0.24) and 0.58 (0.18), respectively. The utility score distributions for the four measures are presented in Figure 1.

Convergent validity

The convergent validity of the four measures is presented in Table 3. We observed significant correlations between all dimension pairs that were hypothesised to measure similar latent traits. Specifically, 15 pairs exhibited strong correlations (range, 0.5–0.77). The associations between QLU-C10D and FACT-8D scores were stronger compared to other comparisons, with correlation coefficients ranging from 0.41 (i.e. role limitation in the QLU-C10D vs. work in the FACT-8D) to 0.77 (nausea in the QLU-C10D vs. nausea in the FACT-8D). In addition, 14 pairs showed moderate correlations (ρ = 0.31–0.49), including seven pairs with EQ-5D-5L scores (ρ = 0.31–0.45), six pairs with SF-6Dv2 scores (ρ = 0.41–0.47) and two pairs between QLU-C10D and FACT-8D (ρ = 0.41 and 0.49, respectively). Moreover, the EQ-5D and SF-6Dv2 utility scores exhibited good agreement (ICC = 0.73). Compared to the other scores, the SF-6Dv2 utility score displayed better agreement with the two cancer-specific measures than the EQ-5D. However, the agreement between the QLU-C10D and FACT-8D scores was poor (r = 0.31). The Bland–Altman plots indicated graphically that the agreement was acceptable for all comparisons with a small mean difference, as few observations were outside the limits of agreement (Figure 2).

Known-group validity

Table 4 presents the known-group validity data for the four measures. Overall, the generic measures (i.e. the EQ-5D and SF-6D) showed higher sensitivity than the cancer-specific

measures (i.e. the QLU-C10D and FACT-8D) in differentiating participants from the different at-risk groups. The EQ-5D was the only measure that could identify differences in HRQoL between all subgroups. While the SF-6D, QLU-C10D and FACT-8D could identify participants diagnosed with cancer stages I or IIA as having a higher HRQoL compared to those diagnosed with cancer stage IIB or greater, the differences were not statistically significant. On the one hand, the F-statistics demonstrated that the EQ-5D had a stronger discriminant power than the other measures in terms of self-care ability, use of assistive tools and cancer stage. On the other hand, the SF-6D was more discriminant in terms of caregiver and treatment statuses.

Discussion

This study is the first to assess the measurement properties of two generic and two cancer-specific PBMs in a group of patients with HL. Overall, the generic measures (i.e. the EQ-5D-5L and SF-6Dv2) showed better measurement performance than the other two cancer-specific measures (i.e. QLU-C10D and FACT-8D). In comparison to the SF-6Dv2, the EQ-5D-5L demonstrated superior convergent and known-group validity, while the SF-6Dv2 exhibited milder ceiling and floor effects when compared to the EQ-5D-5L. These findings demonstrated that the generic PBMs were sensitive enough to measure HRQoL in survivors of HL.

We found that the average utility scores for the SF-6D, QLU-C10D and FACT-8D were lower than that of the EQ-5D-5L and their ceiling effects (range, 9.4%–65.4%) were weaker compared to the EQ-5D-5L (range, 40.8%–91.6%). Previous studies have provided mixed support for these findings. For instance, Gamper et al. demonstrated similar findings when comparing the QLU-C10D with the EQ-5D-3L [15]. Nahvijou et al. showed that the SF-6D generated a higher utility score than the EQ-5D-5L in patients with breast cancer [16]. Kim et al. reported that the QLU-C10D generated a slightly higher utility score than the EQ-5D-3L [17]. Pan et al. found that the QLU-C10D generated a higher mean utility score than the EQ-5D-5L when using five value sets (i.e. Canada, France, Germany, Poland and UK). However, the EQ-5D-5L obtained a higher mean utility score when using three value sets (i.e. Australia, Netherlands and USA) [18]. No comparisons have been reported between the EQ-5D-5L and FACT-8D. There may be several explanations for this finding. For example, the EQ-5D-5L has less dimensions than the SF-6D, QLU-C10D and FACT-8D, which may decrease its utility. Another reason is that the utilities of

the two cancer-specific PBMs were developed using a duration technique combined with a discrete choice experiment (DCE), while most of the EQ-5D-5L utilities were estimated based on a combination of time trade-off (TTO) and DCE techniques. The difference in valuation methods may result in systematic differences in utility. For instance, Xie et al. found that health utility scores derived from DCEs using the duration method were more likely to be lower than values derived using the TTO or DCE techniques [19].

The correlations between the two cancer-specific PBMs were found to be stronger than those between the cancer-specific and generic PBMs. This is not surprising, as both measures were developed to assess the health status specifically for patients with cancer, and their descriptive systems include the most important aspects of HRQoL for this population. For example, the correlation coefficient of nausea between the QLU-C10D and FACT-8D was the highest among all the hypothesised pairs. However, we found that the association between these two utility scores was mild (ICC = 0.31). One possible explanation for this finding is that there is no Chinese value set for the FACT-8D and the Australian preference weights that we used to calculate the utility score may not be appropriate for the Chinese population. There is currently a lack of quantitative evidence for a direct comparison between the QLU-C10D and FACT-8D. A previous study indicated that the PITS state was significantly lower for the FACT-8D than the QLU-C10D, which may result in a lower correlation between the two measures [20]. Another study assessing the content validity of five PBMs in patients with cancer demonstrated that the FACT-8D had the best content validity in terms of relevancy compared with other measures, including the QLU-C10D [21]. In addition, we found a high association between the utility scores of the EQ-5D and SF-6D, but the SF-6D dimensions showed stronger correlations with similar dimensions of the two cancer-specific PBMs than the EQ-5D-5L. Furthermore, the FACT-8D exhibited stronger correlations with the EQ-5D-5L and SF-6D than the QLU-C10D. A previous study provided qualitative evidence that the EQ-5D-5L has good content validity in terms of its comprehensibility [21], but there is a lack of evidence regarding the performance of the SF-6D compared with cancer-specific PBMs, and this requires further evaluation.

Overall, four PBMs demonstrated good discriminant ability in differentiating the HRQoL between at-risk groups, indicating satisfactory known-group validity. The FACT-8D performed

less effectively than the other three PBMs in terms of cancer stage, possibly because the utility score was estimated using Australian preference weights. The F-statistics confirmed that generic PBMs were more sensitive than cancer-specific PBMs in differentiating between patients. On the one hand, we found that the utility scores of generic PBMs were more relevant to daily life functioning, such as self-care, assistive tools and caregivers. On the other hand, the utility score of the QLU-C10D was more relevant for treatment status. This finding contradicted the findings of a previous study by Gamper et al. [15], who found that a more comprehensive descriptive system gives the QLU-C10D a greater advantage in distinguishing between clinically known groups. However, it is worth noting that Gamper et al.'s study had an older sample with more than 80% of the participants reporting a cancer stage of 0 or 1. In contrast, Pan et al.'s study had a sample comparable to ours and found that he EQ-5D-5L utilities generated higher F-statistic values than the corresponding QLU-C10D utilities [18]. Additionally, the variation of the EQ-5D-5L utility score was lower than the variation of the other three measures, which is possibly because it only includes five dimensions, resulting in a milder impact of disease on the overall score [18]. Furthermore, our known-group validity analysis revealed that the EQ-5D-5L consistently showed the largest difference in mean utility between subgroups in three out of five comparisons. For instance, among patients who always used assistive tools, the mean utility score for the EQ-5D-5L was 0.92, while for those who never used them, it was 0.32, resulting in a difference of 0.6. In comparison, the mean differences for the SF-6D, QLU-C10D and FACT-8D were 0.29, 0.3 and 0.19, respectively. This suggests that the utility gain in cost-utility analyses is likely to be larger for the EQ-5D-5L than for the other measures.

Limitations

Several limitations should be considered when interpreting the study findings. First, the sample was recruited from a volunteer pool via a patient organisation's internal network. These volunteers may be patients with better health than most survivors of HL, which may have induced selection bias. Second, all questionnaires were self-reported, which may have led to recall bias. Third, although online surveys are commonly used in this type of research, the data quality may not be entirely guaranteed due to the Internet-based format. Survivors of HL may not be fully engaged in a long survey due to their poor physical and mental health, which may have affected the reliability of our findings. Finally, considering that we collaborated with a

patient association to collect data, some clinical information, such as comorbidities, were not collected based on medical records, which may have affected the validity of our findings.

Conclusions

In conclusion, we found that all four PBMs demonstrated acceptable measurement properties in survivors with HL. Overall, the EQ-5D-5L performed better than the SF-6Dv2 in terms of its agreement with cancer-specific PBMs and discriminant ability. However, the SF-6Dv2 showed stronger associations with similar dimensions of the QLU-C10D and FACT-8D, indicating its high convergent validity. The utility score of the FACT-8D performed poorer than other measures, suggesting the need for the future development of a Chinese value set. In addition, more evidence is urgently needed regarding the performance of the FACT-8D and SF-6Dv2 in those living with cancer. These findings have implications for selecting PBMs to estimate quality-adjusted life years to support economic evaluations for those living with HL.

References

1 Ullah F, Dima D, Omar N, *et al.* Advances in the treatment of Hodgkin lymphoma: Current and future approaches. Front Oncol. 2023;13. https://doi.org/10.3389/fonc.2023.1067289

2 Brazier JE, Yang Y, Tsuchiya A, *et al.* A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. European Journal of Health Economics. 2010;11. https://doi.org/10.1007/s10198-009-0168-z

3 Rowen D, Azzabi Zouraq I, Chevrou-Severac H, *et al.* International Regulations and Recommendations for Utility Data for Health Technology Assessment. Pharmacoeconomics. 2017;35. https://doi.org/10.1007/s40273-017-0544-y

4 Brazier JE, Mulhern BJ, Bjorner JB, *et al.* Developing a New Version of the SF-6D Health State Classification System From the SF-36v2: SF-6Dv2. *Med Care*. 2020;58:557–65.

5 Rowen D, Young T, Brazier J, *et al.* Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. *Value in Health*. 2012;15. doi: 10.1016/j.jval.2012.08.2201

6 King MT, Norman R, Mercieca-Bebber R, *et al.* The Functional Assessment of Cancer Therapy Eight Dimension (FACT-8D), a Multi-Attribute Utility Instrument Derived From

the Cancer-Specific FACT-General (FACT-G) Quality of Life Questionnaire: Development and Australian Value Set. *Value in Health*. 2021;24:862–73.

7 Cella DF, Tulsky DS, Gray G, *et al.* The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *Journal of Clinical Oncology*. 1993;11:570–9.

8 King MT, Costa DSJ, Aaronson NK, *et al.* QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Quality of Life Research*. 2016;25:625–36.

9 Luo N, Liu G, Li M, *et al.* Estimating an EQ-5D-5L Value Set for China. *Value in Health*. 2017;20:662–9.

10 Wu J, Xie S, He X, *et al.* Valuation of SF-6Dv2 Health States in China Using Time Trade-off and Discrete-Choice Experiment with a Duration Dimension. *Pharmacoeconomics*. Published Online First: 2021. doi: 10.1007/s40273-020-00997-1

11 Xu RH, Wong EL, Luo N, *et al.* The EORTC QLU-C10D: the Hong Kong valuation study. *The European Journal of Health Economics*. Published Online First: 2023. doi: 10.1007/s10198-023-01632-4

12 King MT, Norman R, Mercieca-Bebber R, *et al.* The Functional Assessment of Cancer Therapy Eight Dimension (FACT-8D), a Multi-Attribute Utility Instrument Derived From the Cancer-Specific FACT-General (FACT-G) Quality of Life Questionnaire: Development and Australian Value Set. *Value in Health*. 2021;24. doi: 10.1016/j.jval.2021.01.007

13 Gamper EM, Cottone F, Sommer K, *et al.* The EORTC QLU-C10D was more efficient in detecting clinical known group differences in myelodysplastic syndromes than the EQ-5D-3L. *J Clin Epidemiol*. 2021;137. doi: 10.1016/j.jclinepi.2021.03.015

14 Nahvijou A, Safari H, Ameri H. Comparing the performance of the EQ-5D-5L with two versions of the SF-6Dv2 in patients with breast cancer. *Health Serv Outcomes Res Methodol*. 2020;20. doi: 10.1007/s10742-020-00215-7

15 Kim H, Cook G, Goodall S, *et al.* Comparison of EQ-5D-3L with QLU-C10D in Metastatic Melanoma Using Cost-Utility Analysis. *Pharmacoecon Open*. 2021;5. doi: 10.1007/s41669-021-00265-8

16 Pan CW, He JY, Zhu YB, *et al.* Comparison of EQ-5D-5L and EORTC QLU-C10D utilities in gastric cancer patients. *European Journal of Health Economics*. 2023;24. doi: 10.1007/s10198-022-01523-0

17 Xie S, Wu J, He X, *et al.* Do Discrete Choice Experiments Approaches Perform Better Than Time Trade-Off in Eliciting Health State Utilities? Evidence From SF-6Dv2 in China. *Value in Health*. 2020;23. doi: 10.1016/j.jval.2020.06.010

18 McTaggart-Cowan H, King MT, Norman R, *et al.* The FACT-8D, a new cancer-specific utility algorithm based on the Functional Assessment of Cancer Therapies-General (FACT-G): a Canadian valuation study. *Health Qual Life Outcomes*. 2022;20. doi: 10.1186/s12955-022-02002-z

19 Gibson AEJ, Longworth L, Bennett B, *et al.* Assessing the Content Validity of Preference-Based Measures in Cancer. *Value in Health*. 2024;27:70–8.

Table 1 Participant's background characteristics

	N (Mean, SD)	% (Range)
Sex		
Male	259	48.5
Female	275	51.5
Educational level		
Secondary or below	150	28.1
Tertiary or above	384	71.9
Marital status		
Married	319	59.7
Unmarried	215	40.3
Employment status		
Active	242	45.3
Inactive	292	54.7
Caregiver		
No	256	47.9
Yes	278	52.1
Cancer stage		
I	20	3.7
IIA	177	33.1
IIB	68	12.7
III	81	15.2
IV	162	30.3
Unclear	26	4.9
Treatment status		
Being treated	281	52.6
Treatment completed	252	47.2
Age	35.6(11.3)	18~82

Table 2 measurement profile for EQ-5D-5L, SF-6Dv2, QLU-C10D, and FACT-8D

	Level 1		Level 2		Level 3		Level 4		Level 5		Level 6	
	n	%	n	%	n	%	n	%	n	%	n	%
EQ-5D-5L												
Mobility	453	84.8	62	11.6	13	2.4	2	0.4	4	0.7		
Self-care	489	91.6	32	6	6	1.1	-	-	7	1.3		
Usual activities	475	89	44	8.2	10	1.9	2	0.4	3	0.6		
Pain/discomfort	298	55.8	201	37.6	28	5.2	6	1.1	1	0.2		
Anxiety/depression	218	40.8	235	44	63	11.8	10	1.9	8	1.5		
Utility score (SD)	0.89	0.16										
EQ VAS (SD)	78.4	17.4										
SF-6Dv2												
Mental health	116	21.7	161	30.1	208	39	37	6.9	12	2.2		
Pain	274	51.3	121	22.7	104	19.5	26	4.9	6	1.1	3	0.6
Physical functioning	128	24	246	46.1	92	17.2	49	9.2	19	3.6		
Role limitation	131	24.5	133	24.9	188	35.2	58	10.9	24	4.5		
Social functioning	155	29	122	22.8	152	28.5	76	14.2	29	5.4		
Vitality	50	9.4	131	24.5	250	46.8	79	14.8	24	4.5		
Utility score (SD)	0.71	0.19										
QLU-C10D												
Appetite	333	62.4	154	28.8	30	5.6	17	3.2				

Bowel	278	52.1	206	38.6	40	7.5	10	1.9		
Emotional functioning	176	33	254	47.6	64	12	40	7.5		
Fatigue	130	24.3	286	53.6	69	12.9	49	9.2		
Nausea	348	65.2	146	27.3	24	4.5	16	3		
Pain	313	58.6	189	35.4	24	4.5	8	1.5		
Physical functioning	210	39.3	192	36	98	18.4	34	6.4		
Role limitation	349	65.4	137	25.7	27	5.1	21	3.9		
Sleep	240	44.9	227	42.5	38	7.1	29	5.4		
Social functioning	134	25.1	282	52.8	65	12.2	53	9.9		
Utility score (SD)	0.72	0.24								
FACT-8D										
Fatigue	163	30.5	181	33.9	126	23.6	35	6.6	29	5.4
Nausea	306	57.3	112	21	80	15	22	4.1	14	2.6
Pain	269	50.4	166	31.1	79	14.8	11	2.1	9	1.7
Sad	181	33.9	169	31.6	119	22.3	37	6.9	28	5.2
Sleep	96	18	151	28.3	171	32	67	12.5	49	9.2
Support	270	50.6	139	26	85	15.9	34	6.4	6	1.1
Work	138	25.8	119	22.3	127	23.8	76	14.2	74	13.9
Worry	114	21.3	190	35.6	123	23	45	8.4	62	11.6
Utility score (SD)	0.58	0.18								

Table 3 correlations between hypothesized correlations between meaures

No.	Hypothesized correlations			Correlation coefficient
1	EQ-5D MO	~	SF-6D PF	0.44***
2	EQ-5D SC	~	SF-6D SF	0.29***
3	EQ-5D PD	~	SF-6D PA	0.37***
4	EQ-5D AD	~	SF-6D MH	0.7***
5	EQ-5D MO	~	QLU PF	0.42***
6	EQ-5D UA	~	QLU RL	0.44***
7	EQ-5D UA	~	QLU SF	0.31***
8	EQ-5D PD	~	QLU PA	0.58***
9	EQ-5D AD	~	QLU EF	0.61***
10	EQ-5D PD	~	FACT PA	0.6***
11	EQ-5D AD	~	FACT SA	0.55***
12	EQ-5D AD	~	FACT WR	0.45***
13	SF-6D SF	~	FACT WO	0.43***
14	SF-6D PA	~	FACT PA	0.68***
15	SF-6D MH	~	FACT SA	0.57***
16	SF-6D MH	~	FACT WR	0.43***
17	SF-6D VT	~	FACT FA	0.42***
18	SF-6D PF	~	QLU PF	0.47***
19	SF-6D RL	~	QLU RL	0.41***
20	SF-6D SF	~	QLU SF	0.49***
21	SF-6D PA	~	QLU PA	0.6***
22	SF-6D MH	~	QLU EF	0.64***
23	SF-6D VT	~	QLU FA	0.66***
24	QLU RL	~	FACT WO	0.41***
25	QLU EF	~	FACT SA	0.6***
26	QLU EF	~	FACT WR	0.49***
27	QLU PA	~	FACT PA	0.7***

28	QLU FA	~	FACT FA	0.5***
29	QLU SL	~	FACT SL	0.67***
30	QLU NA	~	FACT NA	0.77***

No.	**Hypothesized correlations**			**ICC value**
31	EQ-5D utility	~	SF-6D utility	0.73***
32	EQ-5D utility	~	QLU-C10D utility	0.65***
33	EQ-5D utility	~	FACT-8D utility	0.31***
34	SF-6D utility	~	QLU-C10D utility	0.7***
35	SF-6D utility	~	FACT-8D utility	0.23***
36	QLU-C10D utility	~	FACT-8D utility	0.31***

Table 4 known-group validity

	EQ-5D utility	**SF-6D utility**	**QLU-C10D utility**	**FACT-8D utility**
Self-care ability				
Unable	0.22(0.74)	0.15(0.15)	0.25(0.38)	0.18(0.17)
Little	0.22(0.43)	0.20(0.17)	0.32(0.23)	0.49(0.2)
A lot	0.77(0.18)	0.54(0.19)	0.49(0.27)	0.52(0.23)
Fully capable	0.93(0.09)	0.75(0.16)	0.77(0.19)	0.59(0.17)
F-statistics	322.5	203.1	151.5	20.22
p-value	<0.001	<0.001	<0.001	<0.001
Using assistive tools				
Not at all	0.92(0.11)	0.73(0.17)	0.74 (0.16)	0.59(0.17)
Rare	0.66(0.21)	0.44(0.23)	0.41(0.21)	0.45(0.26)
Sometimes	0.56(0.42)	0.31(0.32)	0.41 (0.11)	0.53(0.18)
Always	0.32(0.72)	0.44(0.38)	0.44 (0.14)	0.4(0.21)
F-statistics	192.2	82.82	49.59	12.28
p-value	<0.001	<0.001	<0.001	<0.001
With a caregiver				
No	0.93(0.08)	0.77(0.16)	0.78(0.19)	0.6(0.16)
Yes	0.86(0.2)	0.66(0.21)	0.66(0.26)	0.56(0.2)
F-statistics	23.13	42.72	38.08	5.69
p-value	<0.001	<0.001	<0.001	0.02
Treatment status				
Treatment completed	0.92(0.12)	0.77(0.16)	0.79(0.19)	0.6(0.15)
Being treated	0.86(0.19)	0.65(0.21)	0.64(0.26)	0.56(0.21)
F-statistics	18.81	56.52	57.83	5.71
p-value	<0.001	<0.001	<0.001	0.02
Cancer stage				
I and IIA (no spread)	0.92(0.11)	0.73(0.18)	0.59(0.16)	0.74(0.23)

≥ IIB (spread)	0.88(0.19)	0.70(0.20)	0.57(0.19)	0.71(0.24)
F-statistics	3.47	2.35	0.40	1.48
p-value	0.03	0.09	0.67	0.23

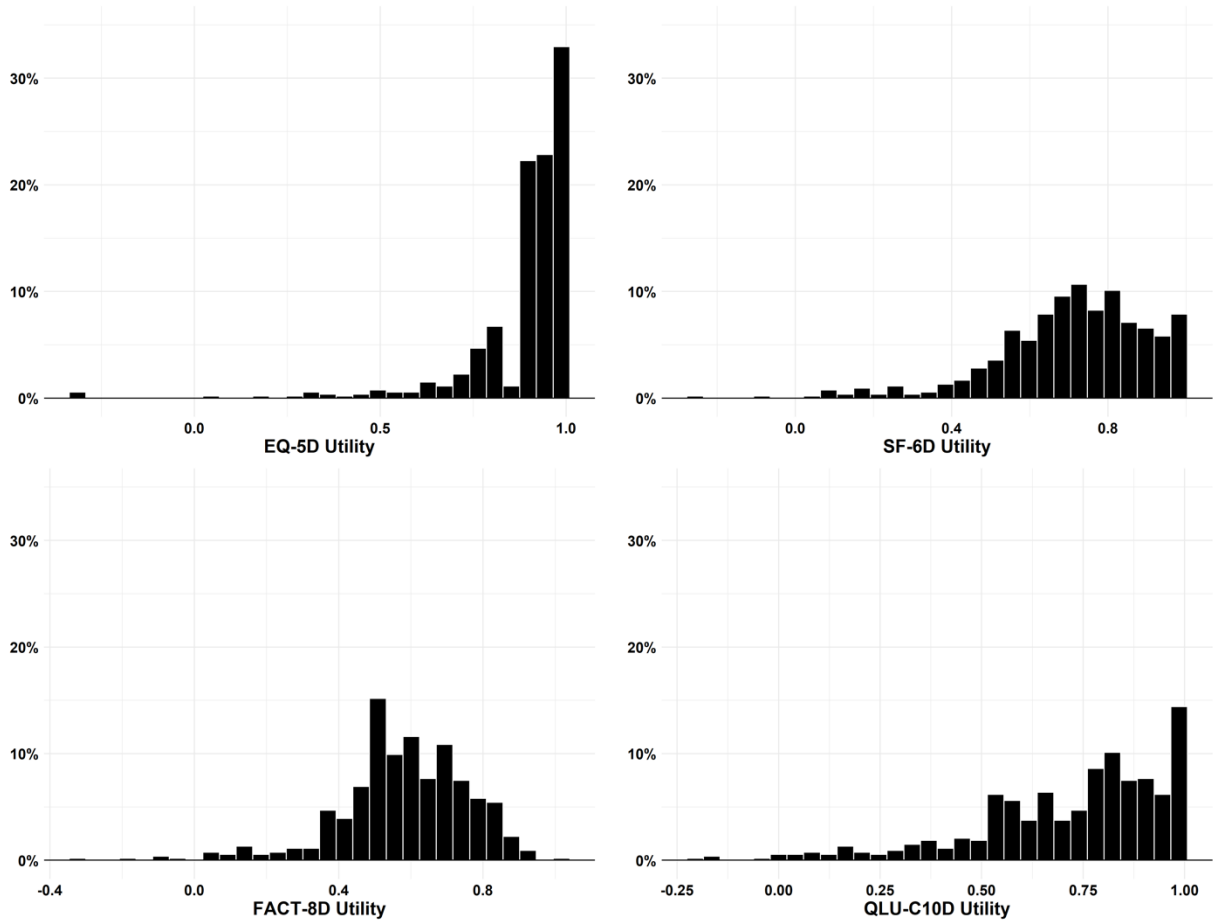

Figure 1 score distribution of four measurements

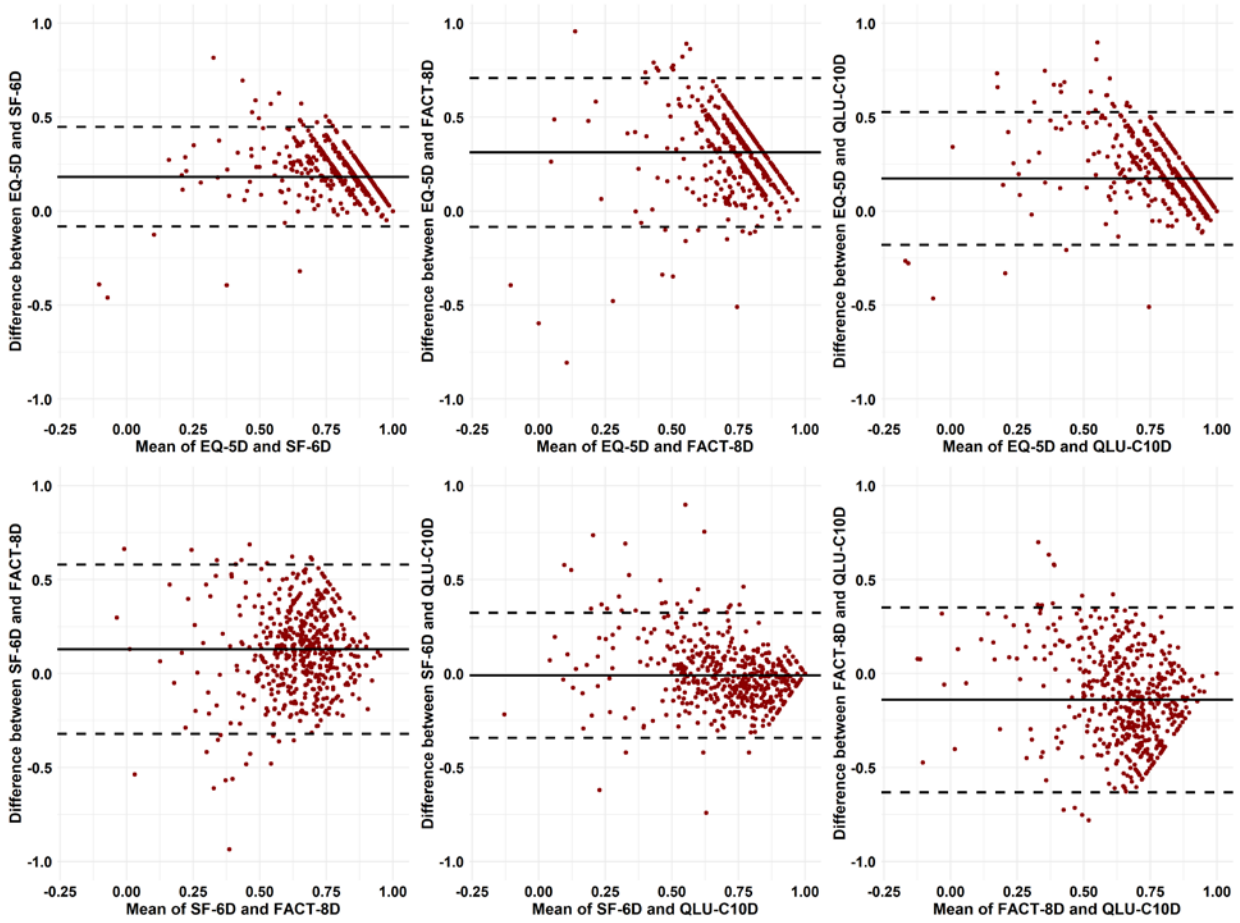

Figure 2 the B-A plot and distribution plot