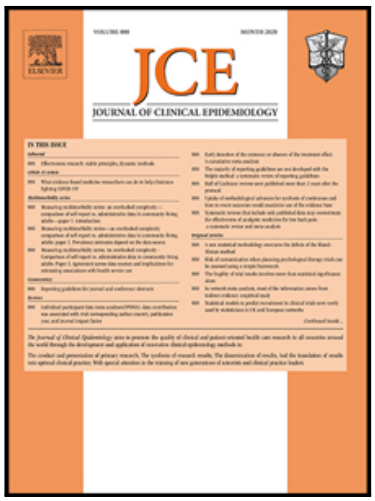

The EORTC QLU-C10D was more efficient in detecting clinical known group differences in myelodysplastic syndromes than the EQ-5D-3L

Eva M Gamper PhD , Francesco Cottone PhD , Kathrin Sommer PhD , Richard Norman PhD , Madeleine King PhD , Massimo Breccia MD , Giovanni Caocci MD , Andrea Patriarca MD , Giuseppe A. Palumbo MD , Reinhard Stauder MD , Pasquale Niscola MD , Uwe Platzbecker MD , Jo Caers MD , Marco Vignetti MD , Fabio Efficace PhD

PII: S0895-4356(21)00088-3
DOI: https://doi.org/10.1016/j.jclinepi.2021.03.015
Reference: JCE 10457

To appear in: *Journal of Clinical Epidemiology*

Accepted date: 15 March 2021

Please cite this article as: Eva M Gamper PhD , Francesco Cottone PhD , Kathrin Sommer PhD , Richard Norman PhD , Madeleine King PhD , Massimo Breccia MD , Giovanni Caocci MD , Andrea Patriarca MD , Giuseppe A. Palumbo MD , Reinhard Stauder MD , Pasquale Niscola MD , Uwe Platzbecker MD , Jo Caers MD , Marco Vignetti MD , Fabio Efficace PhD , The EORTC QLU-C10D was more efficient in detecting clinical known group differences in myelodysplastic syndromes than the EQ-5D-3L, *Journal of Clinical Epidemiology* (2021), doi: https://doi.org/10.1016/j.jclinepi.2021.03.015

MIT: The EORTC QLU-C10D is more efficient in detecting clinical known group differences in myelodysplastic syndromes than the EQ-5D-3L

Eva M Gamper[1], PhD, Francesco Cottone[2], PhD, Kathrin Sommer[2], PhD, Richard Norman[3], PhD, Madeleine King[4], PhD, Massimo Breccia[5], MD, Giovanni Caocci[6], MD, Andrea Patriarca[7], MD, Giuseppe A. Palumbo[8], MD, Reinhard Stauder[9], MD, Pasquale Niscola[10], MD, Uwe Platzbecker[11], MD, Jo Caers[12], MD, Marco Vignetti[2], MD, Fabio Efficace[2], PhD

Affiliations:

[1] Innsbruck Institute of Patient-centered Outcome Research (IIPCOR), Innsbruck, Austria

[2] Italian Group for Adult Hematologic Diseases (GIMEMA), Data Center & Health Outcomes Research Unit, Rome, Italy

[3] School of Public Health, Curtin University, Perth, Australia

[4]University of Sydney, School of Psychology, Sydney, NSW 2006, Australia

[5]Division of Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy.

[6]Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy.

[7]Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy.

[8]Dipartimento di Scienze Mediche, Chirurgiche e Tecnologie Avanzate "G.F. Ingrassia", University of Catania, Catania, Italy

[9]Department of Internal Medicine V (Hematology and Oncology), Innsbruck Medical University, Innsbruck, Austria.

[10]Hematology Unit, Sant'Eugenio Hospital, Rome, Italy

[11]Clinic and Policlinic of Hematology and Cellular Therapy, Hemostaseology, University Hospital Leipzig, Leipzig, Germany

[12]Department of Hematology, CHU de Liège, University of Liege, Liege, Belgium.

Corresponding author:
Eva M Gamper, PhD
Innsbruck Institute of Patient-centered Outcome Research (IIPCOR),
Innsbruck, Austria
Phone: +43 650 7413760; E-mail: eva.gamper@iipcor.org

Abstract

Background: The aim was to investigate the relative validity of the preference-based measure EORTC QLU-C10D in comparison with the EQ-5D-3L in myelodysplastic syndromes (MDS) patients.

Methods: We used data from an international multicentre, observational cohort study of MDS patients. Baseline EORTC QLU-C10D and EQ-5D-3L scores were used and index scores calculated for Italy, Australia, and the UK. Criterion validity was established by Spearman and intraclass correlations (ICC) and Bland-Altman plots. Construct validity was established by the instruments' ability to discriminate known groups, i.e. groups whose health status is expected to differ.

Results: We analysed data from 619 MDS patients (61.1% male; median age 73.8 years). Correlations between theoretically corresponding domains were largely higher than between unrelated domains. ICCs and Bland-Altman plots indicated moderate to good criterion validity. Ceiling effects were lower for the QLU-C10D (4.7%) than for the EQ-5D-3L (22.6%). The EQ-5D-3L failed to discriminate known-groups in two and the QLU-C10D in one of the comparisons; the QLU-C10D's efficiency in doing so was higher in clinical known-groups. Results were comparable between the countries.

Conclusions: The QLU-C10D may be suitable to generate health utilities for economic research in MDS. Responsiveness and minimal important differences need yet to be established.

Keywords: EORTC QLU-C10D, EORTC QLQ-C30, EQ-5D, validity, sensitivity, cancer-specific preference-based measure

Key points

-This is one of the first studies to investigate the sensitivity of the novel cancer-specific preference-based measures QLU-C10D in a clinical setting

- Our results show that cancer-specific health state utility values in a myelodysplastic syndrome population may be determined using the QLU-C10D

- In general, our results inform the ongoing discussion on the arguable advantage of disease-specific over generic preference-based measures

1. Introduction

Myelodysplastic syndrome (MDS) is the umbrella term for are a heterogenous group of haematological neoplasms associated with a high risk of progression to acute myeloid leukaemia (AML) [1]. Age-adjusted incidence rates range between 3.2 and 4.9 [2-5] and increase to 20 per 100.000 in patients aged over 70 [2], clearly making it a disease of the elderly. MDS is highly variable with progression to AML and survival times between a few weeks to several years [6]. Two well-validated disease risk classifications are typically used at diagnostic work up, that is the International Prognostic Scoring System (IPSS) [6] and its more recent revised version, i.e., the IPSS-R [7], that now represents the gold standard in MDS risk classification. These two indices broadly allow to identify lower vs higher risk disease patients and to adopt risk-adapted strategies.

Currently, the only potentially curative approach is stem cell marrow transplantation [8], for which a small percentage of patients is eligible. Other treatments, such as hypomethylating agents or chemotherapy [1, 9] predominantly target increasing time to progression and transfusion dependency as well as symptom relief and health-related quality of life (HRQL).

In MDS research the EQ-5D-3L [10] and the EORTC QLQ-C30 [11] are the most frequently used HRQL outcome measures [12]. The EQ-5D-3L is a preference-based measure and comprises five generic health/HRQL domains, each consisting of one question that is answered by the patient on a 3-level response scale. Hence, it is able to describe $3^5 = 243$ unique health states. A preference-based scoring algorithm allows health states to be valued on a scale between 0 and 1 (where 0 represents being dead and 1 full health). The obtained health state utility values (HSUVs) can be used for the calculation of quality-adjusted life-years (QALYs), which is a core outcome in the economic evaluation of health interventions [13]. Being generic, the EQ-5D-3L is applied across different health conditions and facilitates comparisons between them, which is often required for rational decision making with regard to resource allocation in the health sector [14, 15]. Its

measurement properties, however, vary across conditions. Points of critique include lacking sensitivity in health states with low morbidity [16, 17], inappropriate scale coverage [18-21], and its potential insensitivity to health issues in certain populations of interest due to the limited number and the type of dimensions [22]. A newer version, the EQ-5D-5L [23], overcomes some of the shortcomings of the EQ-5D-3L by using five response levels instead of three [20, 24, 25], but the number of response categories alone does not appear to be decisive [22, 26].Additional approaches have been and are being developed and investigated, such as the SF-6-Dimensions, a generic utility measure based on the SF-36 [27] or the generic PROMIS-Preference score [28]. There is an ongoing debate on the potential added value of disease-specific PBMs which might be able to overcome many of these weaknesses [22].

In contrast to the EQ-5D-3L, the EORTC QLQ-C30 is a HRQL profile measure, not originally designed to provide HSUVs. Its 30 items form 15 scales and cover a broad range of HRQL issues in cancer patients. It can be applied as a research tool as well as for symptom assessment in clinical care. Recently, a preference-based scoring algorithm for the QLQ-C30, the EORTC Quality of Life Utility Core 10 Dimensions (QLU-C10D), has been developed [29, 30]. The QLU-C10D can be used as a cancer-specific preference-based measure as a research tool in health economic research and evaluation. It facilitates calculating utilities from QLQ-C30 data using 13 of its items, which form 10 HRQL domains. With four severity levels per domain, it is able to describe 4^{10}= 1048576 unique health states and therefore may be sensitive to relevant health differences in cancer in general and in MDS in specific.

Generally, it is known that HSUVs differ across instruments [18, 31], but psychometric evidence on the performance of disease-specific utility measures in comparison to generic measures is limited [17]. Disease-specific utility measures are suitable in health technology assessments when a generic preference-based measure is not appropriate or when it shows inferior psychometric performance in a certain condition or patient group [17]. Information from a disease-specific preference-based measure can be useful to improve economic models, e.g. by including them in sensitivity analyses to investigate the potential impact on cost-effectiveness [17, 32].

Because the QLU-C10D is a relatively new measure, this is the first investigation into its psychometric properties. We aimed to examine it in direct comparison to the EQ-5D-3L, using the latter as a preference-based criterion measure with regard to floor and ceiling effects, clinical validity, and relative efficiency in MDS patients. These results will provide first indications whether HSUVs obtained from the QLU-C10D are suitable for use in economic evaluations.

2. Patients and Methods

2.1.Sample/study population and study design

We used baseline data from the PROMYS study, a multicentre, prospective, observational study of adult newly diagnosed MDS patients at 53 centres across 11 countries. Overall, 927 patients were enrolled between 2008 and 2018 and baseline characteristics of the population have been previously published [33]. Enrolment of patients is completed but follow-up of the study is ongoing. The study was initially open to only higher-risk MDS patients and main results on the first cohort of these patients were previously published [34]. The following patient-reported outcome (PRO) measures were included in the study : EORTC QLQ-C30 [35], the FACIT-Fatigue [36], and the Control Preference Scale [37]. In 2014, the study was amended to also consider lower-risk MDS and, amongst other protocol changes, the EQ-5D-3L [10] was also included as an additional PRO measure. Patients had to be diagnosed within 3 months before registration and full details on patient population can be found at ClinicalTrials.gov (NCT00809575). Ethical approval was obtained by the ethical committee of each participating center. Informed consent was received from all participants and the study was performed in accordance with the Helsinki declaration.

2.2.HRQL/PRO measurement

EQ-5D-3L

The EQ-5D-3L descriptive system, which is used to calculate the utility index score, comprises five dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression. Each dimension has three levels indicating “no problems” (level 1), “some problems” (level 2), or “extreme problems” (level 3) [10]. We used the Italian

population tariff for utility calculation [38]. The measurement ranges are as follows: for the Italian version -0.38 to 1.00, for the Australian version -0.217 to 1.00, and for the UK version -0.594 to 1.00. The health state descriptions can be seen in Table A1 in the Appendix.

EORTC QLU-C10D

The EORTC QLU-C10D was derived from the EORTC QLQ-C30 in order to facilitate the calculation of cancer-specific utilities for use in cost-effectiveness analyses. Hence, QLQ-C30 data can be used retrospectively to calculate QLU-C10D utilities. The 13 items selected from the parent instrument form the 10 domains: physical functioning, role functioning, social functioning, emotional functioning, pain, fatigue, nausea, sleep disturbances, bowel problems. Each dimension is described on four severity levels (level 1 = “not at all”, level 2=”a little”, level 3=”quite a bit”, level 4=”very much”). We used the Italian population tariff for utility calculation [39]. The measurement range for the Italian version of the QLU-C10D is 0.025-1.00, for the Australian version -0.096-1.00 and for the UK version -0.083-1.00. The health state descriptions can be seen in Table A2 in the Appendix.

2.3.Statistical methods

Current analysis was based on 619 patients who completed both the EQ-5D-3L and the EORTC QLQ-C30 questionnaire. Sociodemographic characteristics and score distributions were described by proportions, means and standard deviations, median and interquartile ranges, depending on the type of variable. Ceiling and floor effects were assessed by domain and by instrument and were considered relevant when exceeding 15% [40].

Criterion validity was established by different measures of correlations and agreement between the QLU-C10D and the criterion measure EQ-5D-3L. Spearman correlation coefficients were used to evaluate the association between domain and index scores. Correlation coefficients ranging between 0.30-0.49 indicate weak correlations, coefficients

ranging between 0.50-0.69 moderate correlations and correlation coefficients ≥0.70 indicate strong correlations [41]. Given that both measures were designed to assess HRQL, we expected an overall high level of agreement. Yet, since the QLU-C10D includes cancer-specific aspects, some of the instruments' domains were considered to be more closely related than others. We expected higher correlations between theoretically corresponding domains (QLU-C10D physical functioning and EQ-5D mobility, QLU-C10D role functioning and social functioning and EQ-5D usual activities, QLU-C10D pain and EQ-5D pain/discomfort, QLU-C10D emotional functioning and EQ-5D anxiety/depression) and lower correlations between the other domains. Furthermore, responses frequencies on theoretically corresponding domains were cross-tabulated to crudely shed some light on the relation of the different response options. To assess whether utility scores were not merely associated but of the same magnitude, intra-class correlations (ICCs) were calculated as a measure of absolute agreement based on pairs of observations of EQ-5D and QLU-C10D index scores [42]. According to Cachetti (1994) ICC values for absolute agreement between 0.40 and 0.59 indicate moderate agreement values between 0.60 and 0.74 good and values ≥0.75 excellent agreement [42]. A Bland-Altman plot was created to show the extent of agreement between measures, allowing for the identification of the relationship between measurement error and best estimate of the true value [43]. In the lack of minimal important differences (MCIDs) for the Italian EQ-5D-3L index, we used MCIDs reported for the respective UK-based (MCID=0.08) and US-based (MCID=0.06) indexes in an oncology population [44] as a crude measure of clinical importance of disagreement and pre-defined acceptable levels of agreement (LOAs) of +-0.08.The average of both measure was plotted on the x-axis and the difference (QLU-C10D minus EQ-5D) between the measures on the y-axis.

Construct validity was established by investigating the instruments' ability to distinguish between clinical known groups, i.e. groups whose health status is expected to differ. The following known groups were investigated based on what is known on MDS patients from the literature [7, 34, 45]: sex, age (≤73 vs >73) [33], ECOG status (0 vs ≥1), comorbidities measured by the Hematopoietic Cell Transplantation-Comorbidity Index (HTC-CI) [46] (0 vs ≥1 comorbidities), disease risk at diagnosis (i.e., lower vs higher) by the IPSS and IPSS-R indices was considered using the following criteria: IPSS score *lower risk* ="low risk" and "intermediate-1 risk" vs *higher risk* = "intermediate-2" and "high risk"), IPSS-R score

(*lower risk* ="Very low", "Low" and "Intermediate" with an IPSS-R score ≤ 3.5 vs *higher risk*= "High", "Very high" and "Intermediate" with an IPSS-R score > 3.5) [47-49] transfusion dependency (yes vs no) defined as having received at least one red blood cell transfusion every 8 weeks over a period of 4 months [50], haemoglobin (Hb) concentration in (≤10 g/dL vs>10 g/dL). Two known groups of MDS patients were also defined based on the median value of the FACIT Fatigue score in these patients [33], respectively as below or equal to 39 points vs more than 39 points. In additional analyses (data not shown) MDS groups identified by this cut-off showed clinically relevant differences in self-reported symptoms and overall health status/global quality of life measured by the EORTC QLQ-C30 questionnaire [11], according to previously published thresholds [51].

The statistical significance of mean differences between groups was determined with t-tests. As effect size measure we calculated Cohen's d, i.e. the mean difference between groups divided by the pooled standard deviation [52], including 95% confidence intervals. Cohen's guidelines were used to interpret the size of the estimates, i.e. estimates of about 0.2, 0.5 and 0.8 were considered small, medium and large, respectively. To directly compare the instruments with regard to their efficiency to detect differences between groups, the relative efficiency (RE) was calculated as the ratio of F-values [53] with the $F_{EQ\text{-}5D}$ in the denominator and $F_{QLU\text{-}C10D}$ in the numerator. This means that an RE <1 indicates higher efficiency of the EQ-5D-3L and an RE >1 reflects higher efficiency of the QLU-C10D. P-values <.05 were considered statistically significant. All statistical analyses were performed using SAS.

3. Results

3.1.Sample characteristics

Characteristics of the 619 MDS patients analysed are reported in Table 1. Briefly, median age was 73.8 years, the majority (61.1%) were male and mostly diagnosed with lower risk disease according to the IPSS (78.5%).

Insert Table 1

3.2.Floor and ceiling effects

Table 2 and Table 3 show the score frequencies of response levels per domain for the EQ-5D-3L and the QLU-C10D and floor and ceiling effects of the index scores respectively. Between 49.3% and 86.0% reported "no problems" on each of the EQ-5D-3L dimensions, and between 27.5% and 82.1% reported "no problems" for each dimension of the QLU-C10D. Ceiling effects for EQ-5D-3L pain and for EQ-5D-3L anxiety/depression were lower than for the theoretically corresponding domains of the QLU-C10D (pain, emotional functioning) while for the EQ-5D-3L domains mobility and usually activities ceiling effects were higher than for the theoretically corresponding domains of the QLU-C10D (physical functioning, role functioning/social functioning).

With regard to the index scores, the QLU-C10D had lower ceiling effects (4.7%) than the EQ-5D-3L (22.6%), the latter exceeding the 15% that had been defined as relevant.

Insert Tables 2-3

3.3.Criterion validity

3.3.1. Relationship between EQ-5D and QLU-C10D domains

Correlations between the EQ-5D-3L and QLU-C10D domains can be seen in Table 4. As expected, the highest correlation coefficients were found between theoretically related domains. For all three countries these were moderate in size for counterpart domains (ranging between 0.58 and 0.64 for physical functioning and mobility, role functioning and usual activities, and pain and pain/discomfort), except for emotional functioning and anxiety/depression using the Italian tariff where the correlation between these domains was only small (r=0.41). The largest correlation coefficient between a cancer-specific domain and a generic EQ-5D-3L domain was found for fatigue and usual activities (r=0.53). All other correlations coefficients were between r=0.10 (bowel problems and mobility) and r=0.46 (social functioning and usual activities). As expected, correlation coefficients became smaller with increasing content-wise distance of the concepts. The Spearman correlation coefficient between the two index scores was moderate (r=0.70). Cross-tabulations of response frequencies on theoretically corresponding domains indicated especially for the

QLU-C10D physical functioning and the EQ-5D-3L mobility domain and the QLU-C10D emotional functioning and the EQ-5D-3L anxiety/depression domain used different response options on low impairment levels (e.g. reported "no problems" on the EQ-5D-3L mobility item, but "a little problems" when asked for problems making a long walk in on the QLU-C10D physical functioning domain). Details can be seen in see Table A3 in the Appendix.

Insert Table 4

3.3.2. Level of agreement of utility index scores

The ICC were 0.68 (95% CI 0.57- 0.76) for Italy, 0.71 (95% CI 0.67-0.75) for Australia, and 0.67 (95% CI 0.46-0.79) indicating moderate to good absolute agreement between the index scores in all countries. The Bland-Altman plots (see Figure 1) showed the same proportional bias, i.e. the extent of agreement between the measures differed across the measurement range, for all three countries. Overall, the QLU-C10D produced systematically lower scores by -0.0608 in Italy and -0.0808 in the UK. The LOAs showed that in all countries the LOAs exceeded the value of +-0.08 which was defined as acceptable discrepancy.

Insert Figure 1

3.4.Construct validity and relative efficiency

Table 5 shows the sensitivity of both instruments in known group comparisons and their relative efficiency in finding differences between groups. Both instruments detected statistically significant differences between the known groups. Effect sizes were mostly small to moderate. The exceptions were IPSS risk groups, for which neither instrument detected a difference that was statistically significantly different from zero, and IPSS-R risk

groups, in which only the QLU-C10D detected a small difference. The relative efficiency of detecting differences was in favour of the QLU-C10D in the majority of clinical groups in all countries. For the same health states, QLU-C10D scores were systematically lower than EQ-5D-3L scores.

Insert Table 5

4. Discussion

The EORTC QLU-C10D, a novel cancer-specific preference-based measure, has gone through a thorough development process [30, 54-56]. This is the first study that subjected the QLU-C10D to analyses on its criterion and construct validity as it is suggested to be done to examine whether it can be a source of information in economic evaluations [17].

Using the EQ-5D-3L as comparative measure, the results showed that the QLU-C10D has good criterion validity in MDS patients. All analyses were performed pairwise (i.e. using respective national value sets for both measures) for three countries, namely Italy, Australia, and the UK. In all three countries we found lower correlations between theoretically distant domains and higher correlations between theoretically corresponding domains and an overall good association between the index scores. The emotional dimensions showed a small correlation in Italy only. The absolute agreements between index scores were good and the Bland-Altman plots showed proportional bias with LOAs exceeding the size of UK-based and US-based MCIDs of the EQ-5D-3L in oncology [44]. Negative trends, i.e. QLU-C10D scores are lower) were observed.

These results indicate that the two instruments measured a similar construct, but resulting HSUVs are not interchangeable. Therefore, a crucial characteristic of the novel measure concerning its applicability in economic research is its construct validity, i.e. its ability to detect health differences, especially in comparison to established generic measures such as the EQ-5D-3L. This was investigated by performing known group comparisons using both instruments. There was a good agreement with regard to the detection of health/HRQL

differences between groups categorised by health status which were very similar across national value sets. With the exception of the IPSS-R risk groups which was only detected by the QLU-C10D, both measures discriminated the same groups in all three countries. Considering the number of discriminated known groups and the associated efficiency in doing so, the QLU-C10D showed better construct validity than the EQ-5D-3L, especially in clinical groups.

A strong ceiling effect and the weakness in discriminating between good health states has been reported as a frequent issue of concern for the EQ-5D-3L [17]. Investigating the instruments' ceiling effects showed that these were noticeably lower for the QLU-C10D index than for the EQ-5D-3L index; the latter with 22.6% can be considered relevant in size [40]. Response frequencies for the specific domains, however, showed strong ceiling effects for both measures all exceeding the defined threshold of 15%. Focusing on the theoretically corresponding domains revealed lower ceiling effects for the QLU-C10D domains physical functioning, role functioning, and social functioning compared to their theoretical EQ-5D-3L counterparts mobility and usual activities, while they were higher for the QLU-C10D domains pain and depression compared to their theoretical EQ-5D-3L counterparts pain/discomfort and anxiety/depression. One potential explanation for this difference in measurement range might be found in the width of their constructs. This might be especially true for QLU-C10D physical functioning with an item each on the ability to make a long walk and a short walk appears broader than the EQ-5D-3L mobility domain asking for the ability to move around. Vice versa where the EQ-5D-3L exhibited lower ceiling effects the respective domains ask for two aspects each (anxiety and depression; pain and discomfort) while the QLU-C10D asks for only one aspect each (depression; pain). Hence the constructs for these domains captured by the QLU-C10D are narrower and therefore it is "easier" to report no problems. Cross-tabulations of frequencies of responses on these domains showed that indeed the agreement on the two measures on "no problems" was lowest for the QLU-C10D physical functioning "long-walk"- item and EQ-5D-3L mobility and for the QLU-C10D emotional functioning domain and the EQ-5D-3L anxiety/depression domain.

The overall better distribution of the QLU-C10D indices across the measurement range might be results of the higher number of response categories (4 versus 3) and the higher number of included dimensions (10 versus 5). The larger number of dimensions is related to

a larger number of items; 13 used to generate the QLU-C10D utility values versus only 5 used to generate EQ-5D utility values. Of course, this impacts on completion time; the average time to complete the QLQ-C30 is 7-11 minutes [11, 57]; for the EQ-5D a few minutes are given as completion time in their user guide [58]. However, if the QLQ-C30 is used in a trial to assess HRQL as a multi-dimensional construct, as it often is, then the ability to also generate utility scores using the QLU-C10D algorithm saves patients the additional burden of completing a separate preference-based questionnaire, such as the EQ-5D. Further, the richer descriptive system of the QLQ-C30 is likely to be underpin the observed advantage in identifying good health states and in discriminating between clinical known groups, relative to the EQ-5D-3L.

As mentioned earlier, a revised version of the EQ-5D, the EQ-5D-5L [23], aims at overcoming some of the shortcomings of the EQ-5D-3L by using five response levels instead of three [20, 24, 25]. It is a limitation of our study that we cannot provide a direct comparison with this 5-level version. However, so far, there very few value and cross-walk sets available for the EQ-5D-5L and the 3-level version still is the legitimate standard measure, also in the UK where the National Institute for Health and Care Excellence (NICE) has not yet approved the EQ-5D-5L for reimbursement decisions [59]. However, such a comparison will be required to determine the impact of the number of response categories on the relative discriminatory abilities of the measures. There is some evidence that the preference-based precursor of the QLU-C10D, the EORTC-8-Dimension (EORTC-8D) [60] may have greater discriminatory power than the EQ-5D-5L in leukaemia patients [26] and that it better captures changes in HRQL in mild health states in multiple myeloma patients [22]. These findings support that a greater number of response categories alone are not be the sole drivers of differences in sensitivity and efficiency of psychometric instruments.

A question consider is why the QLU-C10D seems to consistently result in lower utilities than the EQ-5D-3L. There are several potential reasons, reflecting key differences between the descriptive systems and the valuation methods. Both these instruments have additive utility algorithms in which each level of each domain that deviates from full health is associated with a utility decrement. The greatest impact is likely to be from the number of domains; the more domains, the greater the summed utility decrement. The impact of the number of levels is less clear; the size of the utility decrement reflects severity, and even

though the EQ-5D only has three levels, the worst level is quite severe in all domains. As Supplementary Table A3 shows, relatively few patients in this study were at the worst levels of the domains on either instrument, so differences in utility decrements at the worst levels are unlikely to be a major contributor to the generally lower values of the QLU-C10D utilities. Table A3 shows that the majority were at the well end of the spectrum, reflecting the fact that MDS patients shortly after diagnosis are generally in relatively good health. The frequencies in Levels 1 and 2 of Table 3 suggest that the QLU-C10D may be able to discriminate impairments at the upper end of the scale more accurately, allowing mild moves away from full health detected by the EQ-5D-3L to contribute to the generally lower values of the QLU-C10D utilities . The final reason why for the QLU-C10D provided systematically lower utility is that the valuations were conducted with a discrete choice experiment, while the EQ-5D valuations used the time-trade-off technique. To our knowledge, whether valuation techniques cause systematic differences and if so, the direction and size of those differences, is a so far unsolved question.

A further limitation of our study is that, since data collection in the PROMYS study is still ongoing, we were unable to provide information on the responsiveness of the QLU-C10D to change. This will be the next step of investigation, along with the question which HRQL dimensions or response characteristics might be driving differences in sensitivity between the measures.

Conclusion

We conclude that the QLU-C10D may be a sensible alternative preference-based measure to generate health utilities for economic research in MDS, if it is not EQ-5D-3L utilities which are required from a regulatory perspective. For example for the National Institute for Care Excellence (NICE) in the UK the EQ-5D-3L still considered the measure of choice in CUAs [61] while the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia does not mandate a specific preference-based measure [62]. In general, there is limited information on HSUVs in MDS; our results provide valuable information on cancer-specific HSUVs in a recently diagnosed MDS population. QALYs can vary substantially depending on baseline evidence, and this variation can even impact health decisions when a cost threshold is in place [63]. Since MDS may progress into AML, HSUVs before progression provide important information for decision analytic models, especially when it is not appropriate to

assume full-health or general population health at baseline. Finally, a big strength of the QLU-C10D is its backward compatibility with the QLQ-C30, which is frequently used in MDS [12]. Hence, in addition to the availability of a 15-scale HRQL profile, HSUVs for MDS can therefore be calculated retrospectively from QLQ-C30 data from a range of existing studies and registry data.

What is new?

-This is the first study to investigate the sensitivity of the novel cancer-specific preference-based measures QLU-C10D in a clinical setting

- Our results show, that cancer-specific health state utility values in a myelodysplastic syndrome population may be determined using the QLU-C10D

- In general, our results inform the ongoing discussion on the arguable advantage of disease-specific over generic preference-based measures

Conflict of interest:
EMG: Received a research grant by the EORTC in 2017 for the investigation of the relative validity of the QLU-C10D in clinical trials. Orcid: 0000-0002-1700-4054
RS was partially supported by Verein Senioren-Krebshilfe. Orcid: 0000-0002-8993-9561
All other authors report no conflict of interest related to this work.

Declarations

Financial Disclosure: None reported.

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material: data may be available from the authors on request

Code availability: code is available on request

References

1. Malcovati L, Hellström-Lindberg E, Bowen D, Adès L, Cermak J, del Cañizo C, Della Porta MG, Fenaux P, Gattermann N, Germing U, Jansen JH, Mittelman M, Mufti G, Platzbecker U, Sanz GF, Selleslag D, Skov-Holm M, Stauder R, Symeonidis A, van de Loosdrecht AA, de Witte T, Cazzola M: **Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet**. In: *Blood. Volume 122*, edn.; 2013: 2943-2964.
2. Rollison DE, Howlader N, Smith MT, Strom SS, Merritt WD, Ries LA, Edwards BK, List AF: **Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs**. *Blood* 2008, **112**(1):45-52.
3. Germing U, Strupp C, Kundgen A, Bowen D, Aul C, Haas R, Gattermann N: **No increase in age-specific incidence of myelodysplastic syndromes**. *Haematologica* 2004, **89**(8):905-910.
4. Avgerinou C, Alamanos Y, Zikos P, Lampropoulou P, Melachrinou M, Labropoulou V, Tavernarakis I, Aktypi A, Kaiafas P, Raptis C, Kouraklis A, Karakantza M, Symeonidis A: **The incidence of myelodysplastic syndromes in Western Greece is increasing**. *Ann Hematol* 2013, **92**(7):877-887.
5. **https://seer.cancer.gov/csr/1975_2017/results_merged/sect_30_mds.pdf#search=mds**
6. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, Sanz M, Vallespi T, Hamblin T, Oscier D, Ohyashiki K, Toyama K, Aul C, Mufti G, Bennett J: **International scoring system for evaluating prognosis in myelodysplastic syndromes**. *Blood* 1997, **89**(6):2079-2088.
7. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, Bennett JM, Bowen D, Fenaux P, Dreyfus F, Kantarjian H, Kuendgen A, Levis A, Malcovati L, Cazzola M, Cermak J, Fonatsch C, Le Beau MM, Slovak ML, Krieger O, Luebbert M, Maciejewski J, Magalhaes SM, Miyazaki Y, Pfeilstocker M, Sekeres M, Sperr WR, Stauder R, Tauro S, Valent P, Vallespi T, van de Loosdrecht AA, Germing U, Haase D: **Revised international prognostic scoring system for myelodysplastic syndromes**. *Blood* 2012, **120**(12):2454-2465.
8. de Witte T, Bowen D, Robin M, Malcovati L, Niederwieser D, Yakoub-Agha I, Mufti GJ, Fenaux P, Sanz G, Martino R, Alessandrino EP, Onida F, Symeonidis A, Passweg J, Kobbe G, Ganser A, Platzbecker U, Finke J, van Gelder M, van de Loosdrecht AA, Ljungman P, Stauder R, Volin L, Deeg HJ, Cutler C, Saber W, Champlin R, Giralt S, Anasetti C, Kröger N: **Allogeneic hematopoietic stem cell transplantation for MDS and CMML: recommendations from an international expert panel**. In: *Blood. Volume 129*, edn.; 2017: 1753-1762.
9. Fenaux P, Haase D, Sanz GF, Santini V, Buske C: **Myelodysplastic syndromes: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up**. *Ann Oncol* 2014, **25 Suppl 3**:iii57-69.
10. **EuroQol--a new facility for the measurement of health-related quality of life**. *Health Policy* 1990, **16**(3):199-208.
11. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, et al.: **The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology**. *J Natl Cancer Inst* 1993, **85**(5):365-376.

12. Stauder R, Lambert J, Desruol-Allardin S, Savre I, Gaugler L, Stojkov I, Siebert U, Chevrou-Severac H: **Patient-reported outcome measures in studies of myelodysplastic syndromes and acute myeloid leukemia: literature review and landscape analysis**. *Eur J Haematol* 2020.
13. Drummond M: **Methods for the Economic Evaluation of Health Care Programmes**, 3 edn. Oxford: Oxford University Press; 2005.
14. Dolan P: **Modeling valuations for EuroQol health states**. *Med Care* 1997, **35**(11):1095-1108.
15. Rowen D, Zouraq IA, Chevrou-Severac H, Hout Bv: **International Regulations and Recommendations for Utility Data for Health Technology Assessment**. *PharmacoEconomics* 2017, **35**(1):11-19.
16. Bharmal M, Thomas J: **Comparing the EQ-5D and the SF-6D Descriptive Systems to Assess Their Ceiling Effects in the US General Population**. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2006, **9**(4).
17. Rowen D, Brazier J, Ara R, Azzabi Zouraq I: **The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment**. *PharmacoEconomics* 2017, **35**(S1):33-41.
18. Brazier J, Roberts J, Tsuchiya A, Busschbach J: **A comparison of the EQ-5D and SF-6D across seven patient groups**. *Health Economics* 2004, **13**(9):873-884.
19. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Loucks L, Al Khudairy A, Noseworthy TW: **Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement**. *Qual Life Res* 2015, **24**(7):1775-1784.
20. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J: **Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study**. *Qual Life Res* 2013, **22**(7):1717-1727.
21. Sullivan PW, Lawrence WF, Ghushchyan V: **A national catalog of preference-based scores for chronic conditions in the United States**. *Med Care* 2005, **43**(7):736-749.
22. Rowen D, Young T, Brazier J, Gaugris S: **Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer**. *Value Health* 2012, **15**(8):1059-1068.
23. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X: **Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L)**. *Qual Life Res* 2011, **20**(10):1727-1736.
24. Kim SH, Kim HJ, Lee SI, Jo MW: **Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea**. *Qual Life Res* 2012, **21**(6):1065-1073.
25. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ: **Comparing the standard EQ-5D three-level system with a five-level version**. *Value Health* 2008, **11**(2):275-284.
26. van Dongen-Leunis A, Redekop W, Uyl-de Groot C: **Which Questionnaire Should Be Used to Measure Quality-of-Life Utilities in Patients With Acute Leukemia? An Evaluation of the Validity and Interpretability of the EQ-5D-5L and Preference-Based Questionnaires Derived From the EORTC QLQ-C30**. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2016, **19**(6).
27. Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36**. *Journal of health economics* 2002, **21**(2):271-292.
28. Hanmer J, Dewitt B, Yu L, Tsevat J, Roberts M, Revicki D, Pilkonis P, Hess R, Hays R, Fischhoff B, Feeny D, Condon D, Cella D: **Cross-sectional validation of the PROMIS-Preference scoring system**. *PloS one* 2018, **13**(7).
29. King MT, Costa DS, Aaronson NK, Brazier JE, Cella DF, Fayers PM, Grimison P, Janda M, Kemmler G, Norman R, Pickard AS, Rowen D, Velikova G, Young TA, Viney R: **QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30**. *Qual Life Res* 2016, **25**(3):625-636.

30. King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, Cella D, Costa DSJ, Fayers PM, Kemmler G, McTaggart-Cowen H, Mercieca-Bebber R, Peacock S, Street DJ, Young TA, Norman R, Consortium M: **Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30**. *Pharmacoeconomics* 2018, **36**(2):225-238.
31. Wee H-L, Machin D, Loke W-C, Li S-C, Cheung Y-B, Luo N, Feeny D, Fong K-Y, Thumboo J: **Assessing Differences in Utility Scores: A Comparison of Four Widely Used Preference-Based Instruments**. *Value in Health* 2007, **10**(4):256-265.
32. Craig D, McDaid C, Fonseca T, Stock C, Duffy S, Woolacott N: **Are Adverse Effects Incorporated in Economic Models? An Initial Review of Current Practice**. *Health technology assessment (Winchester, England)* 2009, **13**(62).
33. Efficace F, Cottone F, Oswald L, Cella D, Patriarca A, Niscola P, Breccia M, Platzbecker U, Palumbo G, Caocci G, Stauder R, Ricco A, Petranovic D, Caers J, Luppi M, Fianchi L, Frairia C, Capodanno I, Follini E, Sarlo C, Fazi P, Vignetti M: **The IPSS-R more accurately captures fatigue severity of newly diagnosed patients with myelodysplastic syndromes compared with the IPSS index**. *Leukemia* 2020.
34. Efficace F, Gaidano G, Breccia M, Voso MT, Cottone F, Angelucci E, Caocci G, Stauder R, Selleslag D, Sprangers M, Platzbecker U, Ricco A, Sanpaolo G, Beyne-Rauzy O, Buccisano F, Palumbo GA, Bowen D, Nguyen K, Niscola P, Vignetti M, Mandelli F: **Prognostic value of self-reported fatigue on overall survival in patients with myelodysplastic syndromes: a multicentre, prospective, observational, cohort study**. *Lancet Oncol* 2015, **16**(15):1506-1514.
35. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC: **The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology**. *J Natl Cancer Inst* 1993, **85**(5):365-376.
36. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E: **Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system**. *Journal of pain and symptom management* 1997, **13**(2):63-74.
37. Degner L, Sloan J, Venkatesh P: **The Control Preferences Scale**. *The Canadian journal of nursing research = Revue canadienne de recherche en sciences infirmieres* 1997, **29**(3).
38. Scalone L, Cortesi PA, Ciampichini R, Belisari A, D'Angiolella LS, Cesana G, Mantovani LG: **Italian population-based values of EQ-5D health states**. *Value Health* 2013, **16**(5):814-822.
39. Gamper E, King M, Norman R, Efficace F, Cottone F, Holzner B, Kemmler G: **EORTC QLU-C10D Value Sets for Austria, Italy, and Poland**. *Quality of life research* 2020, **29**(9):2485-2495.
40. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, Bouter L, de Vet H: **Quality criteria were proposed for measurement properties of health status questionnaires**. *Journal of clinical epidemiology* 2007, **60**(1).
41. Hinkle D, Wiersma W, Jurs S: **Applied Statistics for the Behavioural Sciences**, 5 edn. Boston: Houghton Mifflin; 2003.
42. Cichetti DV: **Guidelines, Criteria, and Rules of Thumb for Evaluating Normed andStandardized Assessment Instruments in Psychology**. *Psychological Assessment,* 1994, **6**(4):284-290.
43. Bland JM, Altman DG: **Measuring agreement in method comparison studies**. *Stat Methods Med Res* 1999, **8**:135-160.
44. S C, M R, P M: **The Minimum Clinically Important Difference for EQ-5D Index: A Critical Review**. *Expert review of pharmacoeconomics & outcomes research* 2014, **14**(2).

45. Efficace F, Cottone F, Abel G, Niscola P, Gaidano G, Bonnetain F, Anota A, Caocci G, Cronin A, Fianchi L, Breccia M, Stauder R, Platzbecker U, Palumbo GA, Luppi M, Invernizzi R, Bergamaschi M, Borin L, Di Tucci AA, Zhang H, Sprangers M, Vignetti M, Mandelli F: **Patient-reported outcomes enhance the survival prediction of traditional disease risk classifications: An international study in patients with myelodysplastic syndromes**. *Cancer* 2018, **124**(6):1251-1259.
46. ML S, MB M, R S, F B, BM S, DG M, B S: **Hematopoietic Cell Transplantation (HCT)-specific Comorbidity Index: A New Tool for Risk Assessment Before Allogeneic HCT**. *Blood* 2005, **106**(8).
47. P F, U P, L A: **How we manage adults with myelodysplastic syndrome**. *British journal of haematology* 2020, **189**(6).
48. Platzbecker U: **Treatment of MDS**. *Blood* 2019, **133**(10).
49. Pfeilstöcker M, Tuechler H, Sanz G, Schanz J, Garcia-Manero G, Solé F, Bennett J, Bowen D, Fenaux P, Dreyfus F, Kantarjian H, Kuendgen A, Malcovati L, Cazzola M, Cermak J, Fonatsch C, Le Beau M, Slovak M, Levis A, Luebbert M, Maciejewski J, Machherndl-Spandl S, Magalhaes S, Miyazaki Y, Sekeres M, Sperr W, Stauder R, Tauro S, Valent P, Vallespi T, van de Loosdrecht A, Germing U, Haase D, Greenberg P: **Time-dependent Changes in Mortality and Transformation Risk in MDS**. *Blood* 2016, **128**(7).
50. L M, U G, A K, MG DP, C P, R I, A G, B H, P B, S K, C S, M L, C A, M C: **Time-dependent Prognostic Scoring System for Predicting Survival and Leukemic Evolution in Myelodysplastic Syndromes**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007, **25**(23).
51. K C, MT K, G V, M MS-J, PM F, JM B: **Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30**. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011, **29**(1).
52. Cohen J: **Statistical Power Analysis for the Behavioural Sciences**, 2 edn: Hillsdale: Lawrence Erlbaum Associates; 1988.
53. Liang MH, Larson MG, Cullen KE, Schwartz JA: **Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research**. *Arthritis & Rheumatism* 1985, **28**(5):542-547.
54. Gamper EM, Holzner B, King MT, Norman R, Viney R, Nerich V, Kemmler G: **Test-Retest Reliability of Discrete Choice Experiment for Valuations of QLU-C10D Health States**. *Value Health* 2018, **21**(8):958-966.
55. Norman R, Kemmler G, Viney R, Pickard AS, Gamper E, Holzner B, Nerich V, King M: **Order of Presentation of Dimensions Does Not Systematically Bias Utility Weights from a Discrete Choice Experiment**. *Value Health* 2016, **19**(8):1033-1038.
56. Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DS, Fayers PM, Kemmler G, Peacock S, Pickard AS, Rowen D, Street DJ, Velikova G, Young TA, King MT: **Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format**. *Qual Life Res* 2016, **25**(3):637-649.
57. Kyriaki M, Eleni T, Efi P, Ourania K, Vassilios S, Lambros V: **The EORTC core quality of life questionnaire (QLQ-C30, version 3.0) in terminally ill cancer patients under palliative care: validity and reliability in a Hellenic sample**. *International journal of cancer* 2001, **94**(1).
58. **EQ-5D-3L User Guide** [Available from: https://euroqol.org/publications/user-guides]
59. **https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l**
60. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, Velikova G: **Deriving a preference-based measure for cancer using the EORTC QLQ-C30**. *Value Health*, **14**(5):721-731.

61. NICE: **National Institute for Health and Care Excellence (NICE): Process and methods guides.The social care guidance manual**. 2013.
62. **Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committe (PBAC) (Version 5.0)** [http://www.pbs.gov.au/industry/listing/elements/pbac-guidelines/PBAC4.3.2.pdf]
63. Ara R, Brazier JE: **Populating an Economic Model with Health State Utility Values: Moving toward Better Practice**. *Value in Health* 2010, **13**(5):509-518.

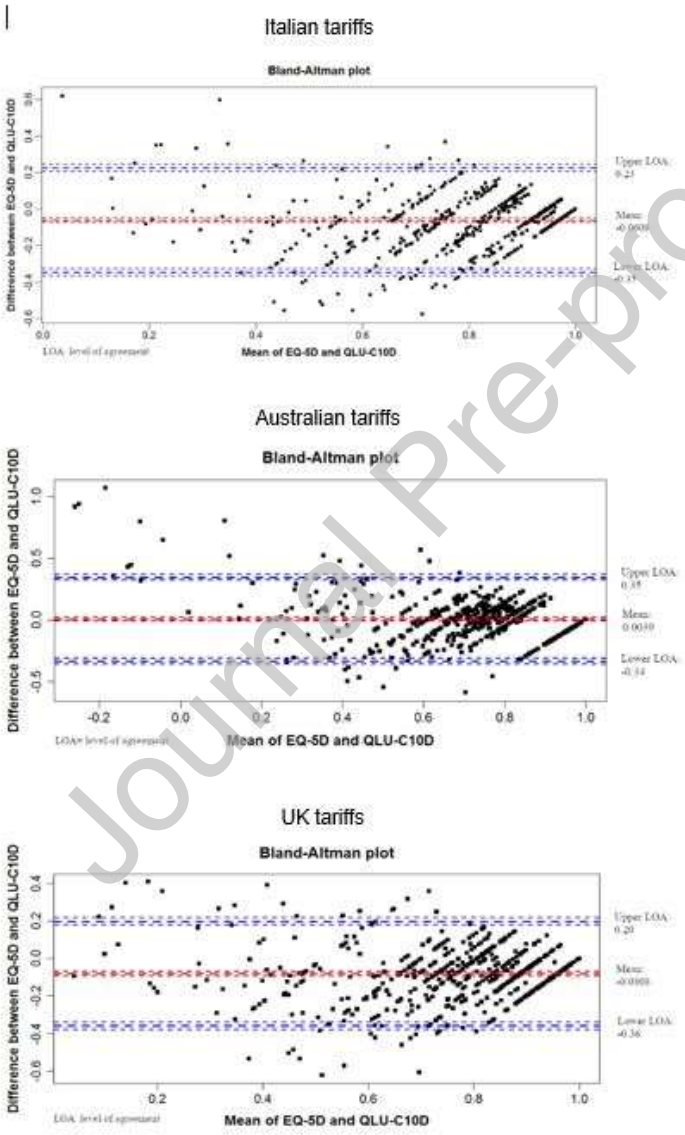

Figure 1: Bland-Altman plots of agreement between the QLU-C10D and EQ-5D-3L.

Table 1: Sociodemographic and clinical information (N=619).

Variable	Value
Age (years)	
Mean (SD)	72.2 (10.7)
Median (IQR)	73.8 (66.8-79.5)
Sex, %	
Male	61.1
Female	38.9
Transfusion dependency[a], %	
No	85.0
Yes	15.0
IPSS[b], %	
Lower risk	78.5
Higher risk	21.5
IPSS-R[c], %	
Lower risk	58.6
Higher risk	41.4
ECOG status, %	
0	46.2
1	42.1
2	10.2
3	1.5
Comorbidities, %	
No	41.9
Yes	58.1
Hb concentration**, g/dL**, %	
<10	46.2
≥10	53.8
FACIT Fatigue score	
Mean (SD)	37.6 (10.8)
Median (IQR)	41.0 (32.0-46.0)
EQ-5D VAS score	
Mean (SD)	63.6 (18.2)
Median (IQR)	65.0 (50.0-80.0)
QLU-C10D index	
Italy:	
Median; Mean (SD)	0.85 ; 0.77 (0.20)
Australia:	
Median; Mean (SD)	0.79; 0.73 (0.20)
UK:	
Median; Mean (SD)	0.79; 0.74 (0.20)
EQ-5D-3L index	
Italy:	
Median; Mean (SD)	0.87; 0.83 (0.19)
Australia:	
Median; Mean (SD)	0.76; 0.73 (0.25)
UK:	
Median; Mean (SD)	0.88; 0.82 (0.19)

SD = standard deviation; IPSS=International Prognostic Scoring System; IPSS-R = International Prognostic Scoring System – Revised; ECOG = Eastern Cooperative Oncology Group; Hb = haemoglobin, FACIT = Functional Assessment of Chronic Illness Therapy; EQ-5D = Euroqol 5 Dimensions; VAS = visual analogue scale; QLU-C10D = Quality of Life Utility Core 10 Dimensions

[a] Transfusion dependency was defined as having received at least one red blood cell transfusion every 8 weeks over a period of 4 months.

[b] "lower risk" category = risk groups "low" and "intermediate-1"; "higher risk" category = risk groups "intermediate-2" and "high".

[c] "lower risk" category = risk groups "very low", "low" and "intermediate" with a IPSS-R score ≤ 3.5; "higher risk" category = risk groups "high", "very high" and "intermediate" with a IPSS-R score > 3.5.

Table 2: Ceiling and floor effects and frequencies of response levels per domain for the EQ-5D-3L

EQ-5D-3L index score	
Ceiling	22.6%
Floor	0.0%

	EQ-5D-3L domains				
	Mobility	**Self-care**	**Usual activities**	**Pain/ Discomfort**	**Anxiety/ Depression**
1 (Ceiling)	358 (57.8%)	532 (86.0%)	371 (60.0%)	326 (52.7%)	305 (49.3%)
2	257 (41.5%)	79 (12.76%)	227 (36.7%)	271 (43.6%)	274 (44.3%)
3 (Floor)	4 (0.65%)	8 (1.29%)	21 (3.4%)	23 (3.7%)	40 (6.5%)

Table 3: Ceiling and floor effects and frequencies of response levels per domain for the QLU-C10D

QLU-C10D index score	
Ceiling	4.7%
Floor	0.0%

	QLU-C10D domains									
	Functioning				**Symptoms**					
	Physical	**Role**	**Social**	**Emotional**	**Pain**	**Fatigue**	**Insomnia**	**Appetite loss**	**Nausea**	**Bowel Problems**
1 (Ceiling)	172 (27.8%)	311 (50.2%)	378 (31.1%)	349 (56.4%)	386 (62.4%)	170 (27.5%)	297 (48.0%)	444 (71.7%)	508 (82.1%)	410 (66.2%)
2	210 (34.0%)	188 (30.4%)	144 (23.3%)	193 (31.2%)	160 (25.9%)	279 (45.1%)	222 (35.9%)	122 (19.7%)	89 (14.4%)	140 (22.6%)

3	159 (25.7 %)	77 (12.4 %)	63 (10.2 %)	54 (8.7%)	55 (8.9%)	123 (19.9 %)	74 (12.0%)	36 (5.8%)	20 (3.2%)	55 (8.9%)
4 (Floor)	78 (12.6 %)	43 (7.0%)	24 (3.9%)	23 (3.7%)	18 (2.9%)	47 (7.6%)	26 (4.2%)	17 (2.8%)	2 (0.3%)	14 (2.3%)

Table 4: Correlations between QLU-C10 and EQ-5D index and domain scores

QLU-C10D Italian tariff	**EQ-5D-3L Italian tariff**				
	index				
index	**0.70***				
			domains		
domains	Mobility	Self-care	Usual activities	Pain/ Discomfort	Anxiety/ Depression
Physical Functioning	**0.58***	0.35*	**0.54***	0.34*	0.25*
Role Functioning	0.43*	0.41*	**0.61***	0.29*	0.32*
Social Functioning	0.25*	0.37*	0.46*	0.20*	0.32*
Emotional Functioning	0.19*	0.22*	0.33*	0.21*	**0.41***
Pain	0.32*	0.21*	0.21*	**0.64***	0.15*
Fatigue	0.39*	0.30*	**0.53***	0.33*	0.35*
Sleep disturbances	0.11*	0.16*	0.24*	0.23*	0.21*
Appetite loss	0.28*	0.34*	0.42*	0.26*	0.28*
Nausea vomitting	0.17*	0.28*	0.28*	0.29*	0.28*
Bowel problems	0.10*	0.11*	0.12*	0.19*	0.15*

QLU-C10D Australian tariff	**EQ-5D-3L Australian tariff**				
	index				
index	**0.72***				
			domains		
domains	Mobility	Self-care	Usual activities	Pain/ Discomfort	Anxiety/ Depression
Physical Functioning	**0.58***	0.35*	**0.54***	0.34*	0.24*
Role Functioning	0.43*	0.41*	**0.61***	0.29*	0.32*
Social Functioning	0.21*	0.38*	0.41*	0.16*	0.25*
Emotional Functioning	0.14*	0.24*	0.35*	0.22*	**0.55***
Pain	0.32*	0.21*	0.21*	**0.64***	0.15*
Fatigue	0.39*	0.30*	**0.53***	0.32*	0.35*
Sleep disturbances	0.11*	0.16*	0.24*	0.23*	0.21*
Appetite loss	0.29*	0.34*	0.43*	0.27*	0.28*
Nausea vomitting	0.17*	0.28*	0.28*	0.29*	0.28*
Bowel problems	0.10*	0.11*	0.12*	0.19*	0.15*
QLU-C10D UK tariff	**EQ-5D-3L UK tariff**				

	index				
index	**0.70***				
	domains				
domains	Mobility	Self-care	Usual activities	Pain/ Discomfort	Anxiety/ Depression
Physical Functioning	**0.58***	0.35*	**0.54***	0.34*	0.245*
Role Functioning	0.43	0.41	**0.61**	0.29	0.32*
Social Functioning	0.25*	0.37*	0.46*	0.20*	0.32*
Emotional Functioning	0.14*	0.24*	0.35*	0.22*	**0.55***
Pain	0.32*	0.21*	0.21*	**0.64***	0.15*
Fatigue	0.39*	0.30*	**0.53***	0.32*	0.35*
Sleep disturbances	0.11*	0.16*	0.24*	0.23	0.22*
Appetite loss	0.28*	0.34*	0.42*	0.26*	0.28*
Nausea vomitting	0.17*	0.28*	0.28*	0.29*	0.28*
Bowel problems	0.10*	0.11*	0.12*	0.19*	0.15*

*p <0.05

Table 5: Sensitivity analyses/known group comparisons and relative efficiency

Tariff		Mean (SD)	Mean (SD)	ES (95% CI)*	p	RE**
		Sex male N=378	**Sex female N=241**			
Italy	QLU-C10D	0.80 (0.19)	0.74 (0.22)	0.30 (0.13-0.46)	≤ 0.001	
Italy	EQ-5D-3L	0.86 (0.16)	0.79 (0.22)	0.38 (0.21-0.54)	≤ 0.001	0.70
Australia	QLU-C10D	0,76 (0,19)	0,69 (0,22)	0.35 (0.18-0.51	≤ 0.001	
Australia	EQ-5D-3L	0,77 (0,20)	0,67 (0,30)	0,41 (0,25-0,57)	≤ 0.001	0,75
UK	QLU-C10D	0,76 (0,18)	0,70 (0,20)	0,32 (0,16-0,48)	≤ 0.001	
UK	EQ-5D-3L	0,84 (0,16)	0,77 (0,23)	0,37 (0,20-0,53)	≤ 0.001	0,76
		Age ≤73 N=284	**Age >73 N=335**			
Italy	QLU-C10D	0.81 (0.19)	0.74 (0.21)	0.35 (0.19-0.51)	≤ 0.001	
Italy	EQ-5D-3L	0.87 (0.16)	0.80 (0.20)	0.38 (0.22-0.54)	≤ 0.001	0.76
Australia	QLU-C10D	0.77 (0.20)	0.70 (0.20)	0.35 (0.19-0.51)	≤ 0.001	
Australia	EQ-5D-3L	0.78 (0.23)	0.69 (0.26)	0.36 (0.20-0.52)	≤ 0.001	0.70

UK	QLU-C10D	0,77 (0,19)	0,70 (0,20)	0,36 (0.20-0.52)	≤ 0.001	
UK	EQ-5D-3L	0,86 (0,16)	0,78 (0,20)	0,44 (0,28-0,60)	≤ 0.001	0,71
		ECOG 0 N=282	**ECOG≥1 N=328**			
Italy	QLU-C10D	0.85 (0.16)	0.71 (0.21)	0.74 (0.58-0.91)	≤ 0.001	
Italy	EQ-5D-3L	0.89 (0.14)	0.79 (0.20)	0.57 (0.41-0.73)	≤ 0.001	1.61
Australia	QLU-C10D	0,81 (0,17)	0,67 (0,21)	0,73 (0,56-0,89)	≤ 0.001	
Australia	EQ-5D-3L	0,80 (0,20)	0,67 (0,28)	0,53 (0,37-0,46)	≤ 0.001	1,61
UK	QLU-C10D	0,81 (0,16)	0,67 (0,20)	0,77 (0,60-0,93)	≤ 0.001	
UK	EQ-5D-3L	0,87 (0,14)	0,77 (0,21)	0,55 (0,39-0,71)	≤ 0.001	1,73
		Comorbidities no N=358	**Comorbidities yes N=258**			
Italy	QLU-C10D	0.82 (0.17)	0.74 (0.22)	0.40 (0.24-0.56)	≤ 0.001	
Italy	EQ-5D-3L	0.89 (0.13)	0.79 (0.21)	0.55 (0.39-0.79)	≤ 0.001	0.60
Australia	QLU-C10D	0,78 (0,17)	0,70 (0,22)	0,40 (0,24-0,56)	≤ 0.001	
Australia	EQ-5D-3L	0,80 (0,17)	0,68 (0,28)	0,54 (0,38-0,70)	≤ 0.001	0,60
UK	QLU-C10D	0,78 (0,17)	0,70 (0,20)	0,44 (0,27-0,60)	≤ 0.001	
UK	EQ-5D-3L	0,87 (0,13)	0,78 (0,21)	0,54 (0,37-0,70)	≤ 0.001	0,62
		Not transfusion dependent[a] N=520	**Transfusion dependent[a] N=92**			
Italy	QLU-C10D	0.78 (0.20)	0.73 (0.20)	0.25 (0.03-0.47)	0.005	
Italy	EQ-5D-3L	0.84 (0.17)	0.78 (0.21)	0.34 (0.12-0.56)	0.005	0.58
Australia	QLU-C10D	0,74 (0,20)	0,69 (0,20)	0.25 (0.03-0.47)	0,006	
Australia	EQ-5D-3L	0,74 (0,24)	0,68 (0,27)	0,27 (0,05-0,49)	0,005	0,58
UK	QLU-C10D	0,74 (0,20)	0,69 (0,19)	0,25 (0,03-0,47)	0,004	
UK	EQ-5D-3L	0,83 (0,18)	0,77 (0,21)	0,32 (0,10-0,55)	0,018	0,84

		IPSS lower risk[b] N=486	IPSS higher risk[b] N=133			
Italy	QLU-C10D	0.78 (0.20)	0.76 (0.21)	0.10 (-0.09-0.29)	0.367	
Italy	EQ-5D-3L	0.83 (0.19)	0.84 (0.18)	-0.05 (-0.24-0.14)	0.278	3.44
Australia	QLU-C10D	0,74 (0,20)	0,72 (0,21)	0,11 (-0,08-0,30)	0,346	
Australia	EQ-5D-3L	0,73 (0,25)	0,74 (0,25)	-0,04 (-0,23-0,15)	0,361	3.44
UK	QLU-C10D	0,74 (0,20)	0,72 (0,20)	0,10 (-0,09-0,29)	0,292	
UK	EQ-5D-3L	0,81 (0,19)	0,83 (0,19)	-0,11 (-0,30-0,09)	0,238	1.32
		IPSS-R lower risk[c] N=363	**IPSS-R higher risk[c] N=256**			
Italy	QLU-C10D	0.79 (0.20)	0.74 (0.21)	0.24 (0.01-0.41)	≤0.001	
Italy	EQ-5D-3L	0.84 (0.18)	0.82 (0.20)	0.11 (-0.01-0.27)	0.492	10.0
Australia	QLU-C10D	0,76 (0,20)	0,71 (0,21)	0,24 (0,008-0,41)	≤0.001	
Australia	EQ-5D-3L	0,74 (0,23)	0,71 (0,28)	0,12 (-0,04-0,28)	0,407	10,0
UK	QLU-C10D	0,76 (0,19)	0,71 (0,20)	0,26 (0,10-0,42)	≤ 0.001	
UK	EQ-5D-3L	0,83 (0,18)	0,81 (0,20)	0,11 (-0,05-0,27)	0,8099	17,3
		Haemoglobin <10 N=286	**Haemoglobin ≥10 N=333**			
Italy	QLU-C10D	0.83 (0.18)	0.72 (0.21)	0.56 (0.39-0.72)	≤ 0.001	
Italy	EQ-5D-3L	0.87 (0.14)	0.80 (0.21)	0.38 (0.22-055)	≤ 0.001	2.09
Australia	QLU-C10D	0,79 (0,18)	0,69 (0,20)	0,52 (0,36-0,68)	≤ 0.001	
Australia	EQ-5D-3L	0,77 (0,21)	0,69 (0,28)	0,32 (0,17-0,48)	≤ 0.001	2,09
UK	QLU-C10D	0,79 (0,18)	0,69 (0,20)	0,52 (0,36-0,68)	≤ 0.001	
UK	EQ-5D-3L	0,85 (0,16)	0,79 (0,21)	0,32 (0,16-0,48)	≤ 0.001	2,79
		FACIT Fatigue <39 N=267	**FACIT Fatigue ≥39 N=349**			
Italy	QLU-C10D	0.62 (0.20)	0.89 (0.11)	-1.73 (-1.91- -1.54)	≤ 0.001	
Italy	EQ-5D-3L	0.71 (0.20)	0.93 (0.09)	-1.48 (-1.66- -1.30)	≤ 0.001	1.41

Australia	QLU-C10D	0,59 (0,20)	0,85 (0,12)	-1,63 (-1,81- -1,44)	≤ 0.001	
Australia	EQ-5D-3L	0,58 (0,28)	0,84 (0,15)	-1,20 (-1,37- -1,03)	≤ 0.001	1,41
UK	QLU-C10D	0,59 (0,19)	0,85 (0,11)	-1,73 (-1,92- -1,54)	≤ 0.001	
UK	EQ-5D-3L	0,70 (0,21)	0,91 (0,11)	-1,30 (-1,48- -1,13)	≤ 0.001	1,65

*ES=effect size Cohen's d

**RE=relative efficiency; ratio of F-statistics; RE>1 QLU is more efficient, RE<1 EQ-5D is more efficient

[a] Transfusion dependency was defined as having received at least one red blood cell transfusion every 8 weeks over a period of 4 months

[b] "lower risk" category = risk groups "low" and "intermediate-1"; "higher risk" category = risk groups "intermediate-2" and "high"

[c] "lower risk" category = risk groups "very low", "low" and "intermediate" with a IPSS-R score ≤ 3.5; "higher risk" category = risk groups "high", "very high" and "intermediate" with a IPSS-R score > 3.5