

RESEARCH

Open Access



EQ-5D-5L and SF-6Dv2 health utilities scores of diffuse large B-cell lymphoma patients in China

Mincai Li¹, Bingxue Fang¹, Hongfei Gu² and Yawen Jiang^{1*}

Abstract

Background This study evaluates the health-related quality of life (HRQoL) of persons with diffuse large B-cell lymphoma (DLBCL) by using EQ-5D-5L and SF-6Dv2 and compares the measurement properties of the two instruments.

Method DLBCL patients were identified via a patient group and were surveyed using web-based questionnaires. Demographic information, socioeconomic status (SES), clinical characteristics, and EQ-5D-5L and SF-6Dv2 responses were collected and statistically described. The association between the EQ-5D-5L and SF-6Dv2 dimensions were analyzed using the Spearman's correlation coefficient, whereas the correlation of the utility scores was evaluated using Pearson's correlation coefficient. The agreement between the responses of the two instruments were examined using a Bland–Altman (B-A) plot. A one-way analysis of variance (ANOVA) was performed to compare the utility scores across subgroups in different clinical states (a t-test was used if there were two subgroups). In addition, the graded response model (GRM) was used to describe the discrimination ability and difficulty characteristics of the dimensions in the two instruments.

Results In total, 582 valid responses were collected, among which 477 respondents were associated with initial-treatment and 105 respondents were relapsed/refractory (RR) patients. The mean (standard deviation [SD]) EQ-5D-5L and SF-6Dv2 utility scores of the DLBCL patients were 0.828 (0.222) and 0.641 (0.220), respectively. The correlation between the EQ-5D-5L and SF-6Dv2 dimensions ranged from 0.299 to 0.680, and the correlation between their utility scores was 0.787. The B-A plot demonstrated an acceptable but not strong agreement between EQ-5D-5L and SF-6Dv2 utility scores. The GRM model results indicated that all dimensions of each instrument were highly discriminating overall, but EQ-5D-5L had suboptimal discriminative power among patients with good health.

Conclusion Both the EQ-5D-5L and SF-6Dv2 showed valid properties to assess the HRQoL of DLBCL patients. However, utility scores derived from the two instruments had substantial difference, thereby prohibiting the interchangeable use of utilities from the two instruments.

Keywords EQ-5D-5L, SF-6Dv2, Diffuse large B-Cell lymphoma, Health-Related quality of life, Measurement properties

*Correspondence:

Yawen Jiang
jiangyw26@mail.sysu.edu.cn

¹ School of Public Health (Shenzhen), Sun Yat-Sen University, Room 533, West Wing of Medical Complex #1, Shenzhen, China

² Hongmian Cancers and Rare Disorders Charity Foundation of Guangzhou, Guangzhou, China

Background

Representing 30 – 40% of all non-Hodgkin lymphoma (NHL) cases, diffuse large B-cell lymphoma (DLBCL) is the most prevalent type of NHL [1]. According to the 2020 GLOBOCAN data, there were an estimated 544,352 new cases and 259,793 deaths of NHL globally, comprising 2.8% of all cancer diagnoses and 2.6% of all cancer deaths [2, 3]. Once developed, DLBCL



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

progresses rapidly and affects tissues and organs such as bone marrow, spleen, thymus, lymph nodes, lymphatic vessels. Currently, the standard of care in the first-line DLBCL treatment is the R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) chemoimmunotherapy regimen and its modifications [4, 5]. Among those who respond to first-line treatment, about 30% to 40% will relapse and 10% will become refractory [6]. Patients who do not respond to second-line therapy or progress after autologous stem cell transplant currently have very limited treatment options [7, 8]. Recently, innovative therapies such as antibody–drug conjugates [9, 10], CAR-T cell therapy [11, 12], and bispecific antibody therapy [13] have been developed to improve the outcomes of DLBCL patients. These innovations underscore the need of assessing their values in the attributes of patient-reported outcomes. However, despite recent progress, there remains a notable evidentiary gap in the health utility values in Chinese DLBCL population.

Overall, DLBCL represents an enormous burden to the patients and society with its many-faceted clinical manifestations, suboptimal treatment responses, and poor prognosis. Therefore, a thorough assessment of its effects on DLBCL patients' HRQoL is imperative.

Generic preference-based measures (GPBMs) are suitable to assess HRQoL across different populations and health conditions [14–16]. They offer comprehensive insights into patient health and overall well-being. In addition, the responses of GPBMs can be converted to utility scores that are necessary components of quality-adjusted life years (QALYs) [17]. The EQ-5D-5L is a GPBM that measures HRQoL based on five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (P/D), and anxiety/depression (AD). The health status of EQ-5D-5L is defined by 5 dimensions, and each of its dimensions has 5 levels [18]. The possible responses to EQ-5D-5L encapsulates a comprehensive range of 3125 health states [19]. The EQ-5D-5L is one of the most popular GPBMs to evaluate health utility scores [20–22]. The reliability and validity of the Chinese versions of EQ-5D-5L have been inspected abundantly, rendering its wide use in China [23]. The SF-6Dv2 is another frequently health-utility-enabled instrument, which was derived from the Short-Form Health Survey (SF-36). The conceptual equivalence between the original English version and the Simplified Chinese version of SF-6Dv2 has been established in previous studies [24–26]. The SF-6Dv2 has six dimensions: physical functioning (PF), role limitation (RL), social functioning (SF), pain (PA), mental health (MH), and vitality (VA). Each of the SF-6Dv2 dimensions other than the PA-dimension has 5 response levels. The PA-dimension, on the other hand, has six

response levels. The SF-6Dv2 descriptive system allows 18,750 health states [25].

To our best knowledge, evidence on the comparative performance of EQ-5D-5L and SF-6Dv2 in Chinese DLBCL patients is currently absent. Therefore, the primary objective of the survey was to describe the HRQoL of DLBCL patients in China using the EQ-5D-5L and SF-6Dv2 questionnaires. The secondary objective of this study was to compare the properties of EQ-5D-5L and SF-6Dv2 among DLBCL patients in China.

Method

Study design and patients

DLBCL patients were recruited via the largest lymphoma patient organization in China. As of 2022, there were more than 100,000 registered members (lymphoma patients-50% and their family members-50%) of the organization. The link to the survey was disseminated by the organization staff via its internal network to the members that registered as DLBCL patients, family members of patients, and caregivers of patients. The survey methodology, execution, and data integrity were defined by the survey committee, including rare disease medical experts, leaders of rare disease patient foundations, and our research team.

To uphold the integrity and quality requirements of data collection, the questionnaire was developed in collaboration with medical experts in DLBCL and blood diseases. The final version of questionnaire was refined through two rounds of pilot testing with a small group of patients to ensure clarity and relevance. The survey was distributed digitally via a secure platform, which ensured structured data collection and minimized the chances of manual data entry errors.

To ensure that each entry in our dataset represented a unique individual, the survey committee employed a manual data exclusion process using Microsoft Excel. Specifically, the survey committee cross-referenced the IP addresses from which the surveys were submitted, the respondents' registered domicile at the city level, the hospital where the diagnosis was made, the disease status, treatment regimen, healthcare expenditures, and the total duration taken to complete the survey. By comparing these data points, duplicate responses were identified and excluded. The survey committee conducted periodic data quality reviews and follow-up contact with participants when necessary to ensure that the data collected were reliable and accurate. This study only included those patients who fully completed the questionnaires and there were no missing values in our analysis.

The main inclusion criteria of this study population included: 1. the patients were adults over 18 years old, 2. the respondent had a diagnosis of DLBCL, or was a

family member or caregiver of a DLBCL patient, and 3. the respondent can describe the patient's disease status, quality of life and other conditions. Before commencing the survey, participants were advised to prepare specific documents that would aid in accurately responding to questions. This preparation ensured that participants had access to comprehensive information regarding the patient's health condition and treatment history, enabling them to meet the inclusion criteria effectively. The main exclusion criteria included: 1. unwilling to report some crucial questions in the survey, 2. received treatment outside mainland China.

Data collection

Demographic and clinical variables

The basic demographic and clinical variables collected were 1) demographic information: sex, age at the time of study, ethnicity, and family registration); 2) SES: marital status, education level, employment status, health insurance coverage, personal income, and household income; and 3) clinical characteristics: treatment lines, whether on treatment, treatment efficacy assessment results, IPI score, double-hit/triple-hit (DHL/THL), double-expressor (DEL), non-GCB subtype, ABC subtype, TP53 mutation, MYD88 mutation and/or CD79b mutation, Ann Arbor staging, and complication.

Treatment stages information

According to a question in our survey, "which of the following is the patient's current status?", patients were classified into five categories: 1, untreated newly diagnosed patients, 2, newly diagnosed patients on treatment, 3, patients completed the first-line treatment and were being monitored for health, 4, RR patients on treatment, and 5, RR patients completed at least one line of treatment after relapse and were being monitored for health. In our analysis, the first three categories were described as initial treatment patients and the rest were described as RR treatment patients.

According to a question in our survey, "what was the efficacy assessment result of the treatment according to the doctor?", patients were classified to four categories: 1, complete remission, 2, partial remission, 3, progressive disease, and 4, have not been evaluated. In this analysis, these answers were described as CR, PR, PD and Unknown.

HRQoL measure instruments

Given the complexity of the functional status of DLBCL patients, especially that of those unable to complete the survey on their own due to severe health conditions or poor digital illiteracy, this study allowed the family members or caregivers to answer on behalf of the

patients [27–31]. This study used the self-reported versions of both EQ-5D-5L and SF-6Dv2 in the questionnaire and the responses were converted to utility values using Chinese-specific value set [26, 32].

Statistical analysis

To ensure the robustness and validity of the statistical analysis, this study performed a Chow's test [33, 34] to evaluate the comparability of the data collected from the self-reported and proxy-reported groups and to determine if the responses of the two groups could be pooled in subsequent analyses [35, 36]. The results of the Chow's test indicated that the null hypothesis could not be rejected, suggesting that pooled analyses of self-reported and proxy-reported data were statistically justified. Detailed results of the Chow's test are provided in Appendix.

Descriptive statistics

Descriptive analyses were conducted for both initial and RR treatment patients. The EQ-5D-5L and SF-6Dv2 utility scores value scores were reported by both treatment stage information and treatment efficacy assessment results. Categorical variables were reported using frequency and percentage. Continuous variables were described using means and standard deviations.

Chi-squared tests were conducted to statistically evaluate the differences in demographic and SES characteristics across initial (untreated, on treatment, and off treatment) and RR (on treatment and off treatment) patients, whereas a t-test was performed to test the difference in age.

The proportion of patients reporting the best and worst levels of the EQ-5D-5L and SF-6Dv2 dimensions and utility values were examined to assess the ceiling and floor effects. Ceiling and floor effects were reported if the proportions of respondents reporting the highest and lowest utility score were greater than 15% [37].

Agreement between the EQ-5D-5L and SF-6Dv2

The agreement between the utility values of the EQ-5D-5L and SF-6Dv2 can be used to determine whether these utility values can be used interchangeably. The Bland–Altman plot is utilized to visually display the agreement of utility values across different ranges. The agreement of utility scores obtained from EQ-5D-5L and SF-6Dv2 was depicted using the B-A plot. The x-axis of the B-A plot represents the mean of the two utility scores whereas the y-axis represents the difference. The mean difference of the two utility scores is labeled using a red horizontal line in the middle of the plot while the upper and lower limits of the range within which 95% of the differences between the two utility scores are expected to lie

are visualized using two black horizontal lines. It is considered “good agreement” if the mean difference is close to zero and more than 95% of the scatters lie within the range [38].

Convergent/Divergent validity

Convergent validity is established when two measures that are expected to be related demonstrate a strong correlation. For instance, the mobility dimension in the EQ-5D-5L is expected to exhibit a significant correlation with the physical functioning dimension in the SF-6Dv2 [39, 40]. Pearson’s correlation coefficients were used to assess the association of utility score (continuous data) and Spearman’s rank correlation coefficients were used to assess the associations between EQ-5D-5L and SF-6Dv2 dimensions (categorical data). Correlations were deemed strong, moderate, and weak when the coefficients were >0.5 , between 0.3 – 0.5 , and <0.3 , respectively [41].

Discriminant validity: known-group validity and GRM

The assessment of known-group validity, a measure that elucidates the capacity of the EQ-5D-5L and SF-6Dv2 to discriminate respondents with different clinical severity, was conducted using one-way ANOVA for multiple groups and the t-test across binary variable groups. Specifically, the groups stratified by treatment lines [42–44], whether on treatment [45], treatment efficacy assessment results [46], IPI score [47, 48], DHL/THL [49, 50], DEL [51, 52], non-GCB subtype, ABC subtype [49], TP53 mutation [51], MYD88 mutation and/or CD79b mutation [53–55], Ann Arbor staging [56, 57], and complication were analyzed [58]. Furthermore, the selection of clinical characteristics defining clinical severity was guided by evidence from previous studies and clinical practice suggesting an association between these characteristics and health status in patients with DLBCL. It was hypothesized that the factors associated with lower utility scores included the RR status, on treatment, PD treatment efficacy assessment results, higher risk IPI score, with DHL/THL, with DEL, non-GCB subtype/ABC subtype, with TP53 mutation, with MYD88 mutation and/or CD79b mutation, higher Ann Arbor staging, and with complications.

Item response theory (IRT) is a statistical model commonly used in psychometric and educational measures. IRT has also been increasingly engaged in quality-of-life research and patient-reported outcomes (PRO) measures in recent years [59–61]. IRT relates a person’s ability/trait level to their probability of answering a question correctly. Based on the findings from previous studies, data collected from EQ-5D-5L and SF-6Dv2 met the key assumptions for the application of IRT [62–64]. The evidence to support the appropriateness of applying IRT

to the current data is provided in Appendix. In the PRO analog, the ability/trait refers to health status. In this study, the GRM, which is an extension of IRT to ordinal response variables, was used to evaluate the discrimination power of each item within each instrument [65].

The GRM generates difficulty (in this study, it represents the underlying health) parameters (b_n) for n response categories, and a discrimination parameter (a) for each dimension [61]. The dimensions correspond to the items of each instrument and the response categories correspond to the levels of each item. The discrimination parameter of an item indicates the capacity of the corresponding dimension to differentiate between study subjects at different health status [60]. A higher value suggests that the item that can more effectively differentiate respondents across the spectrum of health status [61]. To visualize the discrimination parameter and the relationship between a person’s ability and their probability of choosing a less damaged health state, the boundary characteristic curve (BCC) produced from the results of the graded response model GRM can be used. The BCC is a sigmoid shaped curve, typically plotted with ability on the x-axis and the probability of choosing a less damaged health state on the y-axis. On the BCC plot, the discrimination parameter is represented by the slope of the curve. Items with higher discrimination will have steeper slopes. Items with discrimination parameters >0.50 are typically considered being able to adequately discriminate respondents with different ability [61, 65]. The difficulty parameter, alternatively known as the location parameter, indicates the health status where there is a 50% probability of moving to the next higher response category. It is represented by the location on the ability scale where the curve is steepest such that there is maximum uncertainty in whether a respondent will endorse the item. It is widely accepted that the difficulty parameter of an item should fall within the interval of $[-3.00, 3.00]$ to ensure accurate and effective discrimination [59]. To apply GRM to our analysis, the response levels of EQ-5D-5L items and the five-level items of SF-6Dv2 were assigned values of $5 \sim 1$, while the options of the 6-level items of SF-6Dv2 were assigned values of $6 \sim 1$.

Ordinary least square (OLS) linear regression

A multivariable OLS regression was used to examine the factors that affected the EQ-5D-5L and SF-6Dv2 utility scores. Categorical independent variables were reorganized as dummy variables. The level of each category that was supposedly associated with the highest health utility value was employed as the reference category in our regression model. The reference levels of the categorical explanatory variables were initial treatment, off

treatment, CR, IPI=0~1 (low risk), no DHL/THL, no DEL, no TP53 mutation, neither MYD88 mutation nor CD79b mutation, Ann Arbor staging: Stage I, and no complication. The variables that demonstrated statistical significance in the one-way ANOVA were considered for inclusion in the OLS model. The significance level was set at 0.05. All statistical analyses were performed using Stata/MP 15.0 (StataCorp LLC, College Station, TX, USA).

Results

Demographic and SES characteristics

In total, 582 valid responses were collected between September 9 and November 3, 2022. Table 1 shows the demographic and SES characteristics stratified by patients' treatment stages. A total of 477 respondents were associated with initial-treatment patients and 105 respondents were associated with RR patients. Out of the valid responses, 294 were collected from the patients themselves and 288 were provided by the patients' family members or other caregivers. Female patients slightly outnumbered male. With a mean age of 48.95 years, most of the patients with DLBCL in the survey aged 31 to 60 years by the time of the current study. About 97.42% of the patients in the survey were Han Chinese, 70% were urban residents, and 78.87% were married. More than half of the respondents completed their tertiary education. In addition, 20.27% of the patients were unemployed by the time of this study. Nearly all were covered by social health insurance (SHI)—54.47% with urban employee basic medical insurance (UEBMI), 15.81% with urban resident basic medical insurance (URBMI), 18.04% with new rural cooperative medical scheme (NRCMS), and 5.15% with urban resident cooperative medical insurance (URCMI). Nearly one-third (32.65%) of patients had supplemental insurance in addition to basic social health insurance (SHI), while only 1.2% lacked any insurance coverage. Furthermore, 21.99% of respondents reported no personal income for 2021. The mean annual household income across all patients was 114,250 yuan in 2021.

Characteristics of EQ-5D-5L and SF-6Dv2 utility scores

The distribution of responses across each level of the dimensions in EQ-5D-5L and SF-6Dv2 is detailed in Table 2. In the EQ-5D-5L responses, 19.71% of initial treatment patients and 10.48% of the RR patients reported full health (utility score = 1). The utility scores of initial treatment patients ranged from -0.333 to 1, while the scores of RR patients ranged from -0.251 to 1. Additionally, the mean (standard deviation [SD]) SF-6Dv2 score was 0.641 (0.220), with 1.72% of initial treatment and no RR patients reported full health. The utility scores of initial treatment patients ranged from

-0.276 to 1 and RR patients ranged from -0.261 to 0.962. Overall, the mean EQ-5D-5L and SF-6Dv2 scores were higher for initial treatment versus RR patients. Moreover, stronger ceiling effects were observed for the EQ-5D-5L, with 18.04% of total responses indicating full health.

According to the EQ-5D-5L results in Table 2, the self-care dimension had the highest proportion of respondents reporting no problems (85.22%). Following self-care, the dimensions ranked from highest to lowest proportions of no problem responses were: usual activities, mobility, pain/discomfort, and anxiety/depression. Additionally, the proportions reporting no problems were higher among initial versus RR treatment patients across all EQ-5D-5L dimensions.

According to the SF-6Dv2 results in Table 2, the highest proportion of respondents reported no problems on the pain dimension (37.63%). Following pain, the dimensions ranked from highest to lowest proportions of no problem responses were: social functioning, mental health, role limitation, physical functioning, and vitality. Compared to initial treatment patients, RR patients more frequently reported no problems on the pain dimension, but less frequently reported no problems on the other SF-6Dv2 dimensions.

Table 3 shows the means, SDs, 95% confidence interval (CI) ranges, and standard errors (SEs) of the EQ-5D-5L and SF-6Dv2 utility scores stratified by treatment stage and treatment response assessment results. Untreated newly diagnosed patients reported a mean utility of 0.857. Among initial treatment patients on treatment, those assessed as CR, PR, and PD reported utilities of 0.785, 0.688, and 0.472, respectively. For initial treatment patients who completed treatment and were monitored, utilities for those assessed as CR, PR, and PD were 0.883, 0.845, and 0.852, respectively. Among RR patients on re-treatment, utilities for those assessed as CR, PR, and PD were 0.861, 0.658, and 0.796, respectively. For RR patients who completed re-treatment and were monitored, utilities for those assessed as CR, PR, and PD were 0.832, 0.876, and 0.805, respectively.

Agreement between the EQ-5D-5L and SF-6Dv2

The B-A plot (Fig. 1) showed modest agreement between EQ-5D-5L and SF-6Dv2 utility scores. The mean difference was 0.188 (95% CI 0.176 to 0.200). The limits of agreement (LoA) ranged from -0.101 to 0.476, with 93.81% of observations within this interval. Notably, patients in extremely poor health states exhibited greater inconsistency in utility scores between measures. For patients in better health, EQ-5D-5L scores were higher than SF-6Dv2 scores.

Table 1 Demographic and SES information of patients

	Overall		Treatment stages				P-value
			Initial treatment		RR treatment		
	n	(%)	n	(%)	n	(%)	
	582	(100)	477	(81.96)	105	(18.04)	
Demographic information							
Respondent							
Patients	294	(50.52)	246	(51.57)	48	(45.71)	0.277
Family members/caregivers	288	(49.48)	231	(48.43)	57	(54.29)	
Gender							
Female	307	(52.75)	252	(52.83)	55	(52.38)	0.933
Male	275	(47.25)	225	(47.17)	50	(47.62)	
Age group							
18–25	18	(3.09)	13	(2.73)	5	(4.76)	0.114
26–30	43	(7.39)	37	(7.76)	6	(5.71)	
31–40	139	(23.88)	119	(24.95)	20	(19.05)	
41–50	108	(18.56)	90	(18.87)	18	(17.14)	
51–60	132	(22.68)	109	(22.85)	23	(21.90)	
61–65	53	(9.11)	45	(9.43)	8	(7.62)	
66–70	51	(8.76)	40	(8.39)	11	(10.48)	
71–80	33	(5.67)	21	(4.40)	12	(11.43)	
81 and above	5	(0.86)	3	(0.63)	2	(1.90)	
Age-Mean (SD)	48.95	(14.57)	48.39	(14.15)	51.51	(16.16)	0.046
Ethnicity							
Han	567	(97.42)	464	(97.27)	103	(98.10)	0.631
Others	15	(2.58)	13	(2.73)	2	(1.90)	
Family registration							
Urban	406	(69.76)	330	(69.18)	76	(72.38)	0.076
Rural	175	(30.07)	147	(30.82)	28	(26.67)	
Other	1	(0.17)	0		1	(0.95)	
SES information							
Marital status							
Single	62	(10.65)	55	(11.53)	7	(6.67)	0.051
Married	459	(78.87)	367	(76.94)	92	(87.62)	
Divorced/widow(er)	61	(10.48)	55	(11.53)	6	(5.71)	
Education							
Primary or below	51	(8.76)	43	(9.01)	8	(7.62)	0.021
Secondary	235	(40.38)	180	(37.74)	55	(52.38)	
Tertiary or above	296	(50.86)	254	(53.25)	42	(40.00)	
Employment							
Full-time employed	171	(29.38)	160	(33.54)	11	(10.48)	<0.001
Part-time employed	12	(2.06)	11	(2.31)	1	(0.95)	
Take unpaid leave	40	(6.87)	29	(6.08)	11	(10.48)	
Farmers	50	(8.59)	40	(8.39)	10	(9.52)	
Retired	185	(31.79)	144	(30.19)	41	(39.05)	
Students	6	(1.03)	3	(0.63)	3	(2.86)	
Unemployed	118	(20.27)	90	(18.87)	28	(26.67)	
Medical insurance							
Free medical service	58	(9.97)	51	(10.69)	7	(6.67)	0.213
UEBMI	317	(54.47)	267	(55.97)	50	(47.62)	0.120

Table 1 (continued)

	Overall		Treatment stages				P-value
	n	(%)	Initial treatment		RR treatment		
			n	(%)	n	(%)	
	582	(100)	477	(81.96)	105	(18.04)	
URBMI	92	(15.81)	72	(15.09)	20	(19.05)	0.315
NRCMS	105	(18.04)	90	(18.87)	15	(14.29)	0.269
URCMI	30	(5.15)	20	(4.19)	10	(9.52)	0.025
Commercial	132	(22.68)	105	(22.01)	27	(25.71)	0.412
No	7	(1.2)	6	(1.26)	1	(0.95)	0.795
Personal annual income							
No income	128	(21.99)	88	(18.45)	40	(38.10)	0.002
< 5k	25	(4.30)	19	(3.98)	6	(5.71)	
5-10k	33	(5.67)	29	(6.08)	4	(3.81)	
10-30k	95	(16.32)	87	(18.24)	8	(7.62)	
30-50k	108	(18.56)	86	(18.03)	22	(20.95)	
50-100k	127	(21.82)	112	(23.48)	15	(14.29)	
100-200k	50	(8.59)	42	(8.81)	8	(7.62)	
200-300k	8	(1.37)	7	(1.47)	1	(0.95)	
300-500k	7	(1.20)	6	(1.26)	1	(0.95)	
> 500k	1	(0.17)	1	(0.21)	0		
Household annual income							
No income	23	(3.95)	19	(3.98)	4	(3.81)	0.505
< 5k	12	(2.06)	10	(2.10)	2	(1.90)	
5-10k	22	(3.78)	16	(3.35)	6	(5.71)	
10-30k	54	(9.28)	42	(8.81)	12	(11.43)	
30-50k	92	(15.81)	82	(17.19)	10	(9.52)	
50-100k	169	(29.04)	134	(28.09)	35	(33.33)	
100-200k	113	(19.42)	90	(18.87)	23	(21.90)	
200-300k	43	(7.39)	35	(7.34)	8	(7.62)	
300-500k	28	(4.81)	25	(5.24)	3	(2.86)	
> 500k	11	(1.89)	10	(2.10)	1	(0.95)	
Not clear	15	(2.58)	14	(2.94)	1	(0.95)	

RR relapsed/refractory

P-value represents the results of Chi-squared (χ^2) test

Other overseas nationality

UEBMI urban employee basic medical insurance, NRCMS new rural cooperative medical scheme, URBMI urban resident basic medical insurance, URCMI urban resident cooperative medical insurance

Bold value = statistically significant difference at 0.05

Convergent/Divergent validity

The associations between the dimensions of EQ-5D-5L and SF-6Dv2 are presented in Table 4. Spearman’s correlation coefficients demonstrated that all EQ-5D-5L dimensions were significantly correlated with the SF-6Dv2 dimensions. Also, the association between EQ-5D-5L and SF-6Dv2 utility scores was statistically significant and strong (Pearson’s correlation coefficient = 0.787).

Known-group validity: discrimination across different groups

The ANOVA and t-test results in Table 5 demonstrate the known-group validity of the EQ-5D-5L and SF-6Dv2. Both utility measures significantly varied across groups defined by most predetermined clinical characteristics, including treatment status, active treatment, treatment response, IPI score, DEL, TP53 mutation status, MYD88/CD79b mutation status, Ann Arbor stage,

Table 2 Characteristics of EQ-5D-5L and SF-6Dv2 responses

	Overall (n = 582)	Initial treatment (n = 477)	RR treatment (n = 105)
EQ-5D-5L			
Utility score, mean (SD)	0.828 (0.222)	0.841 (0.211)	0.769 (0.260)
Utility score = 1, n (%)	105 (18.04)	94 (19.71)	11 (10.48)
Range of respondent scores	[-0.333,1]	[-0.333,1]	[-0.251,1]
Proportions of responses to dimensions			
MO 1/2/3/4/5, %	73.71/16.32/5.5/2.41/2.06	75.26/16.77/3.77/2.73/1.47	66.67/14.29/13.33/0.95/4.76
SC 1/2/3/4/5, %	85.22/9.28/1.72/1.37/2.41	85.95/9.01/1.89/1.05/2.1	81.9/10.48/0.95/2.86/3.81
UA 1/2/3/4/5, %	77.15/16.49/3.26/1.55/1.55	79.25/15.51/2.52/1.47/1.26	67.62/20.95/6.67/1.9/2.86
P/D 1/2/3/4/5, %	39.35/47.25/10.65/1.72/1.03	41.51/46.33/10.06/1.47/0.63	29.52/51.43/13.33/2.86/2.86
AD 1/2/3/4/5, %	29.38/48.11/17.35/3.78/1.37	29.98/49.27/16.56/3.65/0.63	26.67/42.86/20.95/4.76/4.76
SF-6Dv2			
Utility score, mean (SD)	0.641 (0.220)	0.653 (0.214)	0.586 (0.236)
Utility score = 1, n (%)	10 (1.72)	10 (1.72)	0
Range of respondent scores	[-0.276,1]	[-0.276,1]	[-0.261,0.962]
Proportions of responses to dimensions			
PF 1/2/3/4/5, %	9.79/40.55/25.95/17.87/5.84	10.27/41.30/25.58/17.61/5.24	7.62/37.14/27.62/19.05/8.57
RL 1/2/3/4/5, %	13.57/25.95/36.60/17.53/6.36	14.26/27.67/36.27/16.14/5.66	10.48/18.10/38.10/23.81/9.52
SF 1/2/3/4/5, %	15.81/25.43/29.90/19.93/8.93	17.19/26.21/30.40/17.82/8.39	9.52/21.90/27.62/29.52/11.43
PA 1/2/3/4/5/6, %	37.63/24.91/26.80/8.59/1.03/1.03	37.32/26.42/26.83/7.97/1.05/0.42	39.05/18.10/26.67/11.43/0.95/3.81
MH 1/2/3/4/5, %	14.60/33.85/40.21/9.79/1.55	15.30/35.22/38.78/9.43/1.26	11.43/27.62/46.67/11.43/2.86
VT 1/2/3/4/5, %	6.36/28.18/44.85/17.35/3.26	6.50/29.98/45.70/15.51/2.31	5.71/20.00/40.95/25.71/7.62

MO mobility, SC self-care, UA usual activities, P/D pain/discomfort, AD anxiety/depression, PF physical functioning, RL role limitation, SF social functioning, PA pain, MH mental health, VT vitality

and complications. However, the SF-6Dv2 did not significantly discriminate between groups based on age and DHL/THL subgroups. Furthermore, both instruments exhibited limited discriminative ability across groups defined by Non-GCB subtype, ABC subtype, and complications.

Graded response model

Table 6 presents the estimated item discrimination and difficulty parameters for the EQ-5D-5L and SF-6Dv2 using GRM. All items showed high discrimination. Basically, the item location parameters of the two measures did not present the same patterns. The location parameters of EQ-5D-5L were predominantly concentrated among populations with poorer health levels (most location parameters are negative). This implied that EQ-5D-5L's ability to differentiate among individuals with better health levels was relatively low, a finding that is consistent with its ceiling effect. In contrast, the location parameters of SF-6Dv2 were more evenly distributed, allowing for differentiation across various health levels of the patients. For the Item Information Functions and the BCC graphs of the dimensions EQ-5D-5L and SF-6Dv2, please refer to Appendix.

OLS regressions

Table 7 presents the results of the OLS regressions. The coefficient estimates (b) along with 95% CIs are reported. According to the regression coefficients, EQ-5D-5L utility scores were negatively impacted if patients reported being on treatment (-0.095, 95%CI -0.142, -0.048, $p < 0.001$), PD (-0.104, 95%CI -0.187, -0.021, $p < 0.05$), MYD88 mutation (-0.093, 95%CI -0.175, -0.01, $p < 0.05$), Ann Arbor staging IV (-0.076, 95%CI -0.142, -0.001, $p < 0.05$), and complication (-0.035, 95%CI -0.07, 0, $p < 0.05$).

Similarly, SF-6Dv2 utility scores were negatively impacted if patients reported being on treatment (-0.138, 95%CI -0.182, -0.093, $p < 0.001$), PD (-0.127, 95%CI -0.205, -0.049, $p < 0.001$), having not evaluate the treatment efficacy (-0.07, 95%CI -0.125, -0.015, $p < 0.05$), MYD88 mutation (-0.082, 95%CI -0.16, -0.005, $p < 0.05$), Ann Arbor staging IV (-0.093, 95%CI -0.155, -0.032, $p < 0.01$), and complication (-0.054, 95%CI -0.087, 0.021, $p < 0.05$).

Table 3 EQ-5D-5L and SF-6Dv2 utility scores for health states

	N	%	EQ-5D-5L				F	P-value	SF-6Dv2				F	P-value
			Utility value	SD	Lower range	Upper range			SE	Utility value	SD	Lower range		
Untreated newly diagnosed patients														
Unknown (untreated)	2	0.34	0.857	0.203	-0.967	2.680	0.144	-	0.636	0.290	-1.969	3.241	0.205	-
Newly diagnosed patients on treatment														
CR (on treatment)	36	6.19	0.785	0.142	0.737	0.833	0.024	4.36	0.549	0.155	0.497	0.602	0.026	4.22
PR (on treatment)	15	2.58	0.688	0.364	0.487	0.890	0.094	0.006	0.497	0.195	0.390	0.605	0.050	0.007
PD (on treatment)	11	1.89	0.472	0.357	0.233	0.712	0.108		0.304	0.223	0.155	0.454	0.304	
Unknown (on treatment)	54	9.28	0.759	0.273	0.684	0.833	0.037		0.501	0.223	0.440	0.561	0.030	
Patients completed the first-line treatment and were monitored for health														
CR (off treatment)	305	52.41	0.883	0.176	0.863	0.903	0.010	0.89	0.714	0.191	0.692	0.735	0.011	2.15
PR (off treatment)	34	5.84	0.845	0.126	0.801	0.889	0.022		0.637	0.187	0.702	0.572	0.032	
PD (off treatment)	6	1.03	0.852	0.189	0.654	1.050	0.077		0.657	0.202	0.446	0.869	0.082	
Unknown (off treatment)	14	2.41	0.831	0.186	0.723	0.938	0.050		0.651	0.175	0.550	0.752	0.047	
RR patients on treatment														
CR (on treatment)	11	1.89	0.861	0.152	0.759	0.964	0.046	3.07	0.628	0.160	0.520	0.735	0.048	1.52
PR (on treatment)	12	2.06	0.658	0.246	0.502	0.814	0.071	0.038	0.477	0.295	0.290	0.665	0.085	0.222
PD (on treatment)	7	1.20	0.796	0.201	0.610	0.983	0.076		0.499	0.348	0.177	0.820	0.131	
Unknown (on treatment)	16	2.75	0.535	0.409	0.317	0.753	0.102		0.410	0.251	0.276	0.544	0.063	
RR patients completed at least one line of treatment after relapse and were monitored for health														
CR (off treatment)	37	6.36	0.832	0.191	0.768	0.896	0.031	0.46	0.652	0.169	0.596	0.709	0.028	0.44
PR (off treatment)	14	2.41	0.876	0.148	0.789	0.960	0.040		0.692	0.162	0.598	0.785	0.043	0.727
PD (off treatment)	5	0.86	0.805	0.289	0.446	1.163	0.129		0.593	0.291	0.233	0.954	0.130	
Unknown (off treatment)	3	0.52	0.746	0.145	0.387	1.105	0.083		0.704	0.257	0.067	1.341	0.148	

The F-statistic was the result from ANOVA when comparing the utilities
 Bold value = statistically significant difference at 0.05

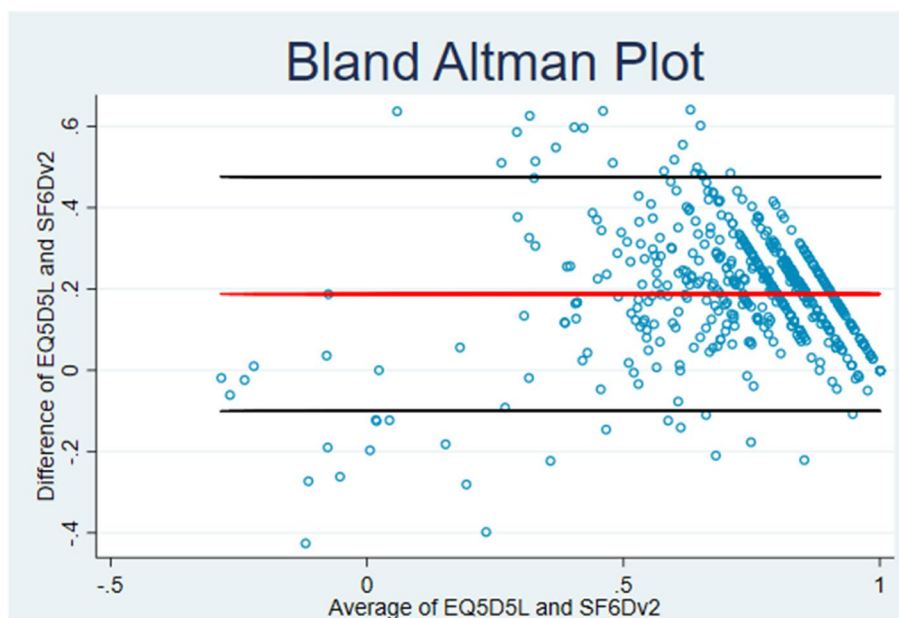


Fig. 1 B-A plot

Table 4 Correlation coefficients of EQ-5D-5L and SF-6Dv2 dimensions and utility scores

	EQ-5D-5L						Utility score
	MO	SC	UA	P/D	AD		
SF-6Dv2							
PF	0.519***	0.466***	0.503***	0.486***	0.336***	-	
RL	0.482***	0.394***	0.497***	0.521***	0.493***	-	
SF	0.501***	0.413***	0.490***	0.497***	0.482***	-	
PA	0.421***	0.359***	0.447***	0.680***	0.387***	-	
MH	0.376***	0.299***	0.374***	0.478***	0.629***	-	
VT	0.410***	0.358***	0.420***	0.536***	0.480***	-	
Utility score	-	-	-	-	-	-	0.787***

MO mobility, SC self-care, UA usual activities, P/D pain/discomfort, AD anxiety/depression, PF physical functioning, RL role limitation, SF social functioning, PA pain, MH mental health, VT vitality

Pearson’s correlation coefficient was provided for utility score (continuous data) and Spearman’s rank correlation coefficient were provided for dimensions (categorical data)

All correlations are significant at the $p < 0.001$ significant level. (* $P < 0.05$, *** $P < 0.01$, **** $P < 0.001$)

Discussion

To our knowledge, this is the first study to compare the measurement properties of the EQ-5D-5L and SF-6Dv2 among DLBCL patients in China. It also provides novel evidence on the health utility values of DLBCL patients across different treatment statuses.

Based on the current findings, DLBCL causes substantial patient quality of life loss, with utility values of patients far below those of the general population. According to literature evidence, the mean utility values of population norms for EQ-5D-5L and SF-6Dv2

are 0.946 and 0.827, respectively [19]. The utility value of EQ-5D-5L for the general population in China is higher than that in the United States [66], but the utility value among DLBCL patients in China is similar to that of the DLBCL patients in the US (alive and responding to Treatment, 0.83) [19, 25, 67, 68]. As such, the DLBCL patients in China may be subject to disproportionately high HRQoL damage. In particular, RR patients and patients with PD were more vulnerable than the other patients, suggesting stronger unmet need in their clinical management.

Table 5 ANOVA and t-test results of clinical characteristics

	n	(%)	EQ-5D-5L				SF-6Dv2			
			Mean	SD	F/t	P-value	Mean	SD	F/t	P-value
Overall	582		0.828	0.222			0.641	0.220		< 0.001
Treatment status										
Initial treatment	477	(81.96)	0.841	0.211	3.02	0.003	0.653	0.214	2.81	0.005
RR treatment	105	(18.04)	0.769	0.260			0.586	0.236		
Whether on treatment										
Untreated	2	(0.34)	0.857	0.203	30.66	< 0.001	0.636	0.290	58.69	0.037
On treatment	162	(27.84)	0.718	0.286			0.496	0.226		
Off treatment	418	(71.82)	0.871	0.175			0.697	0.190		
Treatment evaluation										
CR	389	(66.84)	0.869	0.176	15.69	< 0.001	0.690	0.191	25.42	< 0.001
PR	75	(12.89)	0.789	0.229			0.594	0.217		
PD	29	(4.98)	0.686	0.319			0.474	0.290		
Have not been evaluated	89	(15.29)	0.731	0.298			0.518	0.233		
IPI score										
IPI=0~1 (low risk)	162	(36.16)	0.884	0.126	13.99	< 0.001	0.691	0.168	12.62	< 0.001
IPI=2 (medium risk)	114	(25.45)	0.851	0.178			0.671	0.211		
IPI=3~5 (high risk)	172	(38.39)	0.770	0.268			0.582	0.243		
^a 134 respondents were Unknown										
Double-hit/Triple-hit (DHL/THL)										
Yes	49	(11.26)	0.775	0.247	2.41	0.017	0.620	0.210	1.25	0.213
No	386	(88.74)	0.849	0.195			0.660	0.200		
^a 147 respondents were Unknown										
Double-expressor (DEL)										
Yes	166	(38.88)	0.799	0.254	3.52	< 0.001	0.598	0.237	3.96	< 0.001
No	261	(61.12)	0.868	0.153			0.679	0.179		
^a 155 respondents were Unknown										
Non-GCB subtype/ABC subtype										
Yes	300	(72.64)	0.820	0.227	0.95	0.344	0.628	0.220	1.26	0.207
No	113	(27.36)	0.843	0.207			0.658	0.201		
^a 169 respondents were Unknown										
TP53 mutation										
Yes	76	(23.17)	0.783	0.290	2.59	0.010	0.598	0.255	2.61	0.009
No	252	(76.83)	0.857	0.192			0.669	0.194		
^a 254 respondents were Unknown										
MYD88 mutation and/or CD79b mutation										
CD79b mutation	13	(5.39)	0.794	0.226	4.96	0.002	0.623	0.223	5.1	0.002
MYD88 mutation	31	(12.86)	0.705	0.310			0.517	0.259		
Both	34	(14.11)	0.794	0.276			0.615	0.226		
Neither	163	(67.63)	0.866	0.192			0.674	0.197		
^a 341 respondents were Unknown										
Ann Arbor staging										
Stage I	74	(14.92)	0.906	0.110	9.73	< 0.001	0.729	0.141	9.3	< 0.001
Stage II	118	(23.79)	0.867	0.132			0.666	0.170		
Stage III	87	(17.54)	0.862	0.177			0.665	0.200		
Stage IV	217	(43.75)	0.776	0.276			0.589	0.258		
^a 86 respondents were Unknown										
Complication										
Have at least one complication	240	(41.24)	0.851	0.205	3.02	0.003	0.597	0.232	4.08	< 0.001

Table 5 (continued)

	n	(%)	EQ-5D-5L				SF-6Dv2			
			Mean	SD	F/t	P-value	Mean	SD	F/t	P-value
Have no complication	342	(58.76)	0.795	0.241			0.671	0.206		

The test statistic provided a t-value for the comparison of two groups and the F-statistic is provided from ANOVA when comparing more than two groups

The p-value in the upper right corner represents the significant level of the means of the two scales in the overall patients, and is written in italics

Bold value = statistically significant difference at 0.05

CR complete remission, PR partial remission, PD progressive disease, Unknown respondents whose progression has not been evaluated

^a Unknown respondents of IPI score, DHL/THL, DEL, Non-GCB subtype/ABC subtype, TP53 mutation, MYD88 mutation and/or CD79b mutation, Ann Arbor staging was not included in ANOVA or t-test analysis

Table 6 GRM item parameter estimates

	Discrimination parameter	Difficult parameter				
	a	b1	b2	b3	b4	b5
EQ-5D-5L						
MO	4.478	-2.274	-1.853	-1.366	-0.678	
SC	3.897	-2.257	-2.006	-1.751	-1.125	
UA	8.147	-2.3	-1.986	-1.539	-0.745	
P/D	2.252	-3.029	-2.458	-1.365	0.339	
AD	1.193	-4.136	-2.909	-1.287	0.934	
Mean (SD)	3.993(2.382)	-2.799(0.729)	-2.242(0.391)	-1.462(0.167)	-0.255(0.767)	
SF-6Dv2						
PF	1.811	-2.208	-0.984	0.002	1.797	
RL	3.518	-1.715	-0.822	0.291	1.234	
SF	3.264	-1.544	-0.674	0.244	1.144	
PA	1.81	-3.352	-2.91	-1.711	-0.404	0.455
MH	2.55	-2.607	-1.452	0.05	1.294	
VA	2.903	-2.142	-0.963	0.46	1.807	
Mean (SD)	2.643(0.66)	-2.261(0.597)	-1.301(0.758)	-0.111(0.732)	1.145(0.741)	-

The findings showed systematically higher average utility scores for the EQ-5D-5L versus the SF-6Dv2, aligning with previous research in other cancer populations [19]. The results also revealed a ceiling effect in the EQ-5D-5L but not in the SF-6Dv2 among the DLBCL population (18.04% versus 1.72%). No floor effect was observed for either measure. Previous studies among general population have also shown that the EQ-5D-5L exhibits a greater ceiling effect compared to the SF-6Dv2 [69–71]. Within the realm of cancer research, a study focusing on Chinese patients with breast cancer reported a ceiling effect of 28.6% for the EQ-5D-5L [72]. In the current sample of DLBCL patients, a higher EQ-5D-5L ceiling effect was also observed. This finding is consistent with prior research conducted in both general and disease-specific populations [69, 73, 74]. The difference in the ceiling effects between the EQ-5D-5L and the SF-6Dv2 may be partially attributed to the differing recall periods used by the two measures. Specifically, the SF-6Dv2

frames its questions in the context of health "over the past 4 weeks," whereas the EQ-5D-5L refers to the health status "today." The longer recall period in SF-6Dv2 could potentially allow respondents to consider minor impairments impacting HRQoL that might be overlooked within a shorter timeframe [75].

In addition, this study found a strong correlation between the utility values of EQ-5D-5L and SF-6Dv2, and all dimensions' Spearman correlation coefficients were significantly positive. This is similar to previous research results [71, 75]. This indicates that EQ-5D-5L and SF-6Dv2 have certain similarities in detecting the trend of health utility value changes in DLBCL patients. However, there may be differences in measuring the absolute value of HRQoL, which may be attributable to the different dimensions covered by the two tools. The larger number of dimensions covered by SF-6Dv2 also allows more descriptive states than EQ-5D-5L, thereby engendering greater granularity of health status.

Table 7 OLS regression of clinical characteristics on EQ-5D-5L and SF-6Dv2

	EQ-5D-5L				SF-6Dv2			
	Coef	(95% CI)			Coef	(95% CI)		
Treatment status								
RR treatment	-0.017	-0.064	0.029		-0.005	-0.048	0.039	
Whether on treatment								
Untreated	0.004	-0.29	0.298		-0.035	-0.312	0.242	
On treatment	-0.095	-0.142	-0.048	***	-0.138	-0.182	-0.093	***
Treatment evaluation								
PR	-0.033	-0.088	0.021		-0.046	-0.097	0.005	
PD	-0.104	-0.187	-0.021	*	-0.127	-0.205	-0.049	***
Have not been evaluated	-0.055	-0.114	0.004		-0.07	-0.125	-0.015	*
IPI score								
IPI=2 (medium risk)	0.012	-0.044	0.067		0.03	-0.022	0.082	
IPI=3~5 (high risk)	-0.013	-0.07	0.045		0.009	-0.045	0.063	
Unknown	0.001	-0.056	0.054		0.004	-0.048	0.056	
Double-hit/Triple-hit (DHL/THL)								
Yes	-0.042	-0.108	0.025		-	-	-	
Unknown	-0.012	-0.08	0.056		-	-	-	
Double-expressor (DEL)								
Yes	-0.012	-0.058	0.034		-0.031	-0.072	0.009	
Unknown	-0.031	-0.1	0.038		-0.021	-0.065	0.023	
TP53 mutation								
Yes	-0.017	-0.073	0.039		-0.005	-0.057	0.048	
Unknown	-0.02	-0.068	0.028		-0.018	-0.063	0.027	
MYD88 mutation and/or CD79b mutation								
CD79b mutation	-0.004	-0.124	0.117		0.02	-0.093	0.132	
MYD88 mutation	-0.093	-0.175	-0.01	*	-0.082	-0.16	-0.005	*
Both	-0.025	-0.104	0.053		-0.009	-0.083	0.065	
Unknown	0.01	-0.039	0.06		0.011	-0.035	0.057	
Ann Arbor staging								
Stage II	-0.018	-0.08	0.045		-0.043	-0.102	0.016	
Stage III	-0.003	-0.074	0.069		-0.027	-0.094	0.04	
Stage IV	-0.076	-0.142	-0.011	*	-0.093	-0.155	-0.032	**
Unknown	0.034	-0.107	0.04		-0.033	-0.102	0.036	
Complication								
Have at least one complication	-0.035	-0.07	0	*	-0.054	-0.087	-0.021	***
Constant	0.959	0.904	1.014	***	0.789	0.737	0.84	***

Reference category: initial treatment, off treatment, CR, IPI = 0~1 (low risk), no DHL/THL, no DEL, no TP53 mutation, neither MYD88 mutation nor CD79b mutation, Ann Arbor staging: Stage I, no complication

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Both tools exhibited robust discriminative capabilities across subgroups with different clinical characteristics. While both measures were able to distinguish between groups of disparate severity, the ANOVA analysis revealed some differences in the absolute utility values of EQ-5D-5L and SF-6Dv2 across these subgroups, SF-6Dv2 was able to differentiate more known groups with statistically significant differences compared to EQ-5D-5L, which suggests that the known-group validity of

SF-6Dv2 may be slightly superior to that of EQ-5D-5L among this specific group of patients, a finding that aligns with previous research [69, 74].

The inter-subgroup differences measured by EQ-5D-5L and SF-6Dv2 are relatively similar and much smaller than the absolute differences across the two instruments. For example, the utility values of EQ-5D-5L and SF-6Dv2 for patients undergoing initial treatment were higher by 0.072 and 0.067, respectively,

compared to those for patients experiencing relapse. Similarly, the utility values of EQ-5D-5L and SF-6Dv2 for patients under follow-up observation were higher by 0.153 and 0.201, respectively, compared to those for patients currently in treatment. Furthermore, the utility values of EQ-5D-5L and SF-6Dv2 for patients in CR were higher by 0.183 and 0.216, respectively, compared to those for patients with PD. This suggests that while these two tools may differ in measuring the utility value of a particular health state, they are relatively consistent in reflecting changes or improvements brought about by interventions.

The strengths of this study are manifold. Firstly, they study included both newly treated patients and those who were refractory or had relapsed, allowing stratified analyses by their recurrence status. Secondly, the study benefits from a relatively substantial sample size, encompassing 31 provinces and autonomous regions in China, thereby engendering good representativeness. Lastly, the utilization of the GRM in the analysis provided a novel perspective on the discriminative capacities of the two instruments.

Several limitations of the study must be noted. Firstly, the survey was conducted online, which might have skewed the overall patient age towards a younger demographic, given that older individuals tend to be less proficient with internet usage. Secondly, during the survey process, three scales were used in a fixed order: EQ-5D-5L, QLQ-C30, and SF-6Dv2. This non-random sequence could potentially influence the results. For instance, after responding to questions in the QLQ-C30, patients might reflect more deeply on their affliction, leading them to report a poorer health status when subsequently filling out the SF-6Dv2.

Despite the measures taken to ensure the accuracy and reliability of the information, it should be noted that the current study did not engage objective metrics to determine whether proxy reporters can accurately reflect patient condition, which represents a limitation of our study. Therefore, our conclusions should be interpreted with caution.

Proxy assessments can be obtained by asking proxies to either estimate the patient's responses (i.e., proxy-patient perspective) or provide their own view of the patient's HRQoL (i.e., proxy-proxy perspective) [27]. In practical applications, this survey highly possibly encompassed both perspectives. Furthermore, it is plausible that the patients themselves described their own health states, with proxies merely assisting in operating the smartphone. This is particularly relevant given that cognitive impairment, which could necessitate more assessments from the proxy-proxy perspective, is uncommon in DLBCL.

Conclusions

In a nutshell, DLBCL patients in China experience impaired health utilities, which were especially low among the patients with progressive and relapsed diseases.

In addition, the results of this study demonstrate that both EQ-5D-5L and SF-6Dv2 possess certain degrees of validity and reliability in assessing the HRQoL of patients with DLBCL. However, the measurement outcomes of the two tools exhibit differences, indicating that they cannot be used interchangeably. The findings underscore the importance of selecting an appropriate measurement tool to accurately assess the HRQoL of DLBCL patients, the consideration of which should also hinge on study objectives.

Abbreviations

95% CI	95% Confidence interval
AD	Anxiety/depression
B-A plot	Bland-Altman plot
BCC	Boundary characteristic curve
DEL	Double-expressor
DHL/THL	Double-hit/triple-hit
DLBCL	Diffuse large B-cell lymphoma
GPBMs	Generic preference-based measures
GRM	Graded response model
HRQoL	Health-related quality of life
HTA	Health technology assessment
IRT	Item response theory
LoA	Limits of agreement
MH	Mental health
MO	Mobility
NHL	Non-Hodgkin lymphoma
NRCMS	New rural cooperative medical scheme
OLS	Ordinary least square
P/D	Pain/discomfort
PA	Pain
PF	Physical functioning
PRO	Patient-reported outcomes
QALYs	Quality-adjusted life years
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone
RL	Role limitation
RR	Relapsed/refractory
SC	Self-care
SD	Standard deviation
SES	Socioeconomic status
SEs	Standard errors
SF	Social functioning
SHI	Social health insurance
UA	Usual activities
UEBMI	Urban employee basic medical insurance
URBMI	Urban resident basic medical insurance
URCMI	Urban resident cooperative medical insurance
VT	Vitality

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12955-024-02297-0>.

Supplementary Material 1. See Appendix.

Acknowledgements

We thank the House086 for assisting in questionnaire design and data collection, and the DLBCL patients and their family members/caregivers for completing the questionnaires.

Authors' contributions

ML, HG, and YJ were responsible for the study design; ML, BF and HG were responsible for the recruitment of respondents and data collection. ML and YJ were responsible for data cleaning and statistical analysis. ML and YJ were responsible for the drafting of the manuscript. YJ was responsible for obtaining the funding. All authors read and approved the final manuscript.

Funding

This study was supported by the Hongmian Cancers and Rare Disorders Charity Foundation of Guangzhou.

Availability of data and materials

The data are available from the lead and corresponding authors upon reasonable requests.

Data Availability

The data that support the findings of this study are available from the lead and corresponding authors upon reasonable requests, but restrictions apply to the availability of these data due to privacy.

Declarations

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Sun Yat-sen University School of Public Health (Shenzhen) (No.:2022056).

Consent for publication

All participants were required to provide informed consents before answering the questionnaire. Participants were informed about their freedom of refusal and they could withdraw from the survey at any point. Identifying information of the respondents was not collected.

Competing interests

The authors declare no competing interests.

Received: 7 February 2024 Accepted: 9 September 2024

Published online: 19 September 2024

References

- Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. *Pathology*. 2018;50:74–87.
- GLOBOCAN. <https://gco.iarc.fr/today/fact-sheets-cancers>. Accessed 31 May 2023.
- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–49.
- Guidelines for patients-Diffuse Large B-Cell Lymphomas, Version 2022. NCCN guideline. <https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1480>. Accessed 28 May 2023.
- Poeschel V, Held G, Ziepert M, et al. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *Lancet*. 2019;394:2271–81.
- Ernst M, Oeser A, Besiroglu B, et al. Chimeric antigen receptor (CAR) T-cell therapy for people with relapsed or refractory diffuse large B-cell lymphoma. *Cochrane Database Syst Rev*. 2021;9:CD013365.
- Kanas G, Ge W, Quek RGW, et al. Epidemiology of diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) in the United States and Western Europe: population-level projections for 2020–2025. *Leuk Lymphoma*. 2022;63:54–63.
- Hutchings M. The evolving therapy of DLBCL: Bispecific antibodies. *Hematol Oncol*. 2023;41:107–11.
- Abrisqueta P, Gonzalez-Barca E, Panizo C, et al. Polatuzumab vedotin plus rituximab and lenalidomide in patients with relapsed or refractory diffuse large B-cell lymphoma: a cohort of a multicentre, single-arm, phase 1b/2 study. *Lancet Haematol*. 2024;11:e136–46.
- Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. *N Engl J Med*. 2022;386:351–63.
- Awasthi R, Pacaud L, Waldron E, et al. Tisagenlecleucel cellular kinetics, dose, and immunogenicity in relation to clinical factors in relapsed/refractory DLBCL. *Blood Adv*. 2020;4:560–72.
- Thudium Mueller K, Grupp SA, Maude SL, et al. Tisagenlecleucel immunogenicity in relapsed/refractory acute lymphoblastic leukemia and diffuse large B-cell lymphoma. *Blood Adv*. 2021;5:4980–91.
- Dickinson MJ, Carlo-Stella C, Morschhauser F, et al. Glofitamab for relapsed or refractory diffuse large B-cell lymphoma. *N Engl J Med*. 2022;387:2220–31.
- Kaplan RM, Hays RD. Health-related quality of life measurement in public health. *Annu Rev Public Health*. 2022;43:355–73.
- Brazier J, Ara R, Rowen D, et al. A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics*. 2017;35:21–31.
- Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2018;19:557–70.
- 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. *Eur J Health Econ*. 2010;11:215–25.
- EuroQol Research Foundation. EQ-5D-5L User Guide, Version 3.0, 2019. <https://euroqol.org/publications/user-guides>. Accessed 16 Aug 2022.
- Xie S, Wu J, Xie F. Population norms for SF-6Dv2 and EQ-5D-5L in China. *Appl Health Econ Health Policy*. 2022;20:573–85.
- Xu RH, Wong ELY, Jin J, et al. Mapping of the EORTC QLQ-C30 to EQ-5D-5L index in patients with lymphomas. *Eur J Health Econ*. 2020;21:1363–73.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727–36.
- Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21:1245–57.
- Luo N, Li M, Liu GG, et al. Developing the Chinese version of the new 5-level EQ-5D descriptive system: the response scaling approach. *Qual Life Res*. 2013;22:885–90.
- Cheung PWH, Wong CKH, Samartzis D, et al. Psychometric validation of the EuroQol 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. *Scoliosis Spinal Disord*. 2016;11:19.
- Wu J, Xie S, He X, et al. The simplified Chinese version of SF-6Dv2: translation, cross-cultural adaptation and preliminary psychometric testing. *Qual Life Res*. 2020;29:1385–91.
- 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*. 2021;39:521–35.
- Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care*. 2005;43:493–9.
- Logsdon RG, Gibbons LE, McCurry SM, et al. Assessing quality of life in older adults with cognitive impairment. *Psychosom Med*. 2002;64:510–9.
- Neumann PJ, Araki SS, Gutterman EM. The use of proxy respondents in studies of older adults: lessons, challenges, and opportunities. *J Am Geriatr Soc*. 2000;48:1646–54.
- Higginson IJ, Gao W, Jackson D, et al. Short-form Zarit caregiver burden interviews were valid in advanced conditions. *J Clin Epidemiol*. 2010;63:535–42.
- Magaziner J, Simonsick EM, Kashner TM, et al. Patient-proxy response comparability on measures of patient health and functional status. *J Clin Epidemiol*. 1988;41:1065–74.
- Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L Value Set for China. *Value Health*. 2017;20:662–9.

33. Sundaram V, Jalan R, Shah P, et al. Acute on chronic liver failure from nonalcoholic fatty liver disease: a growing and aging cohort with rising mortality. *Hepatology*. 2021;73:1932–44.
34. Chiang JK, Lai NS, Wang MH, et al. A proposed prognostic 7-day survival formula for patients with terminal cancer. *BMC Public Health*. 2009;9:365.
35. Martins C, Webster EK, Romo-Perez V, et al. Sex differences in 3- to 5-year-old children's motor competence: A pooled cross-sectional analysis of 6241 children. *Scand J Med Sci Sports*. 2024;34:e14651.
36. Dabbour E, Dabbour O, Martinez AA. Temporal stability of the factors related to the severity of drivers' injuries in rear-end collisions. *Accid Anal Prev*. 2020;142:105562.
37. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293–307.
38. Selva-Sevilla C, Ferrara P, Geronimo-Pardo M. Interchangeability of the EQ-5D and the SF-6D, and comparison of their psychometric properties in a spinal postoperative Spanish population. *Eur J Health Econ*. 2020;21:649–62.
39. Heslin M, Chua KC, Trevillion K, et al. Psychometric properties of the five-level EuroQoL-5 dimension and Short Form-6 dimension measures of health-related quality of life in a population of pregnant women with depression. *BJPsych Open*. 2019;5:e88.
40. Suarez L, Tay B, Abdullah F. Psychometric properties of the World Health Organization WHOQOL-BREF Quality of Life assessment in Singapore. *Qual Life Res*. 2018;27:2945–52.
41. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen*. 2012;141:2–18.
42. Friedberg JW. Relapsed/refractory diffuse large B-cell lymphoma. *Hematology Am Soc Hematol Educ Program*. 2011;2011:498–505.
43. Schuster SJ, Bishop MR, Tam CS, et al. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma. *N Engl J Med*. 2019;380:45–56.
44. Colasia A, Njue A, Trask PC, et al. Clinical efficacy and safety in relapsed/refractory diffuse large B-cell lymphoma: a systematic literature review. *Clin Lymphoma Myeloma Leuk*. 2014;14:343–55 e6.
45. Van Den Neste E, Schmitz N, Mounier N, et al. Outcome of patients with relapsed diffuse large B-cell lymphoma who fail second-line salvage regimens in the international CORAL study. *Bone Marrow Transplant*. 2016;51:51–7.
46. Schuster SJ, Tam CS, Borchmann P, et al. Long-term clinical outcomes of tisagenlecleucel in patients with relapsed or refractory aggressive B-cell lymphomas (JULIET): a multicentre, open-label, single-arm, phase 2 study. *Lancet Oncol*. 2021;22:1403–15.
47. Eertink JJ, Zwezerijnen GJC, Heymans MW, et al. Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. *Blood*. 2023;141:3055–64.
48. Eertink JJ, van de Brug T, Wiegers SE, et al. (18)F-FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2022;49:932–42.
49. Nowakowski GS, Czuczman MS. ABC, GCB, and double-hit diffuse large B-cell lymphoma: does subtype make a difference in therapy selection? *Am Soc Clin Oncol Educ Book*. 2015;32:e449–57.
50. Friedberg JW. How I treat double-hit lymphoma. *Blood*. 2017;130:590–6.
51. Doderio A, Guidetti A, Marino F, et al. Dose-adjusted EPOCH and rituximab for the treatment of double expressor and double-hit diffuse large B-cell lymphoma: impact of TP53 mutations on clinical outcome. *Haematologica*. 2022;107:1153–62.
52. Roh J, Cho H, Pak HK, et al. BCL2 super-expressor diffuse large B-cell lymphoma: a distinct subgroup associated with poor prognosis. *Mod Pathol*. 2022;35:480–8.
53. Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378:1396–407.
54. Chen R, Zhou D, Wang L, et al. MYD88(L265P) and CD79B double mutations type (MCD type) of diffuse large B-cell lymphoma: mechanism, clinical characteristics, and targeted therapy. *Ther Adv Hematol*. 2022;13:20406207211072840.
55. Visco C, Tanasi I, Quaglia FM, et al. Oncogenic mutations of MYD88 and CD79B in diffuse large B-Cell lymphoma and implications for clinical practice. *Cancers (Basel)*. 2020;12:2913.
56. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol*. 2014;32:3059–68.
57. Liu X, Cao D, Liu H, et al. Clinical features analysis and survival nomogram of primary small intestinal diffuse large B-cell lymphoma. *Cancer Manag Res*. 2022;14:2639–48.
58. Kinoshita T, Watanabe T, Itoh K, et al. Clinical characteristics of patients with B-cell lymphoma enrolled in clinical trials for aggressive lymphoma in Japan: Japan Clinical Oncology Group - Lymphoma Study Group study - JCOG0108A. *J Clin Exp Hematop*. 2021;61:35–41.
59. Jiang Y, Hesser JE. Using item response theory to analyze the relationship between health-related quality of life and health risk factors. *Prev Chronic Dis*. 2009;6:A30.
60. Fryback DG, Palta M, Cherepanov D, et al. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. *Med Decis Making*. 2010;30:5–15.
61. Reise SP, Du H, Wong EF, et al. Matching IRT models to patient-reported outcomes constructs: the graded response and log-logistic models for scaling depression. *Psychometrika*. 2021;86:800–24.
62. Liu TH, Ho AD, Hsu YT, et al. Validation of the EQ-5D in Taiwan using item response theory. *BMC Public Health*. 2021;21:2305.
63. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38:1128–42.
64. Kopec JA, Sayre EC, Davis AM, et al. Assessment of health-related quality of life in arthritis: conceptualization and development of five item banks using item response theory. *Health Qual Life Outcomes*. 2006;4:33.
65. Kraska J, Bell K, Costello S. Graded response model analysis and computer adaptive test simulation of the depression anxiety stress scale 21: evaluation and validation study. *J Med Internet Res*. 2023;25:e45334.
66. Jiang R, Janssen MFB, Pickard AS. US population norms for the EQ-5D-5L and comparison of norms from face-to-face and online samples. *Qual Life Res*. 2021;30:803–16.
67. Chen Q, Staton AD, Ayer T, et al. Exploring the potential cost-effectiveness of precision medicine treatment strategies for diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2018;59:1700–9.
68. Chimeric antigen receptor T-cell therapy for B-cell cancers: effectiveness and value. ICER (Institute for Clinical and Economic Review); 2018.
69. Kangwanrattanukul K. A comparison of measurement properties between UK SF-6D and English EQ-5D-5L and Thai EQ-5D-5L value sets in general Thai population. *Expert Rev Pharmacoecon Outcomes Res*. 2021;21:765–74.
70. Yang Z, Busschbach J, Liu G, et al. EQ-5D-5L norms for the urban Chinese population in China. *Health Qual Life Outcomes*. 2018;16:210.
71. Xie S, Wang D, Wu J, et al. Comparison of the measurement properties of SF-6Dv2 and EQ-5D-5L in a Chinese population health survey. *Health Qual Life Outcomes*. 2022;20:96.
72. Li S, Wang M, Liu L, et al. Which approach is better in eliciting health state utilities from breast cancer patients? Evidence from mainland China. *Eur J Cancer Care (Engl)*. 2019;28:e12965.
73. Sun CY, Liu Y, Zhou LR, et al. Comparison of EuroQoL-5D-3L and short form-6D utility scores in family caregivers of colorectal cancer patients: a cross-sectional survey in China. *Front Public Health*. 2021;9:742332.
74. Ye Z, Sun L, Wang Q. A head-to-head comparison of EQ-5D-5 L and SF-6D in Chinese patients with low back pain. *Health Qual Life Outcomes*. 2019;17:57.
75. Xu RH, Dong D, Luo N, et al. Evaluating the psychometric properties of the EQ-5D-5L and SF-6D among patients with haemophilia. *Eur J Health Econ*. 2021;22:547–57.

Publisher's Note

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