Towards patients' understanding of health-related quality of life—a mixed-method study in psoriasis and multiple sclerosis

Valerie Andrees[1] · Joachim Westenhöfer[2] · Christine Blome[1] · Christoph Heesen[3,4] · Matthias Augustin[1] · Janine Topp[1]

Abstract

Purpose Patients' individual understanding of health-related quality of life (HRQoL) varies widely, making the measurement of this complex and subjective construct challenging. Anchoring vignettes, i.e., descriptions of fictive patients may provide insights into patients' individual questionnaire reference frames, assessment processes, and understanding of HRQoL. This study analyzes how patients assess HRQoL of vignettes.

Methods This exploratory mixed-method study included 100 patients with a chronic disease (50 multiple sclerosis (MS); 50 psoriasis). Sixteen vignettes, two for each domain of the SF-12v2, were developed based on literature recommendations and pretested in a convenience sample of seven healthy individuals. Patients assessed their own HRQoL and HRQoL of the vignettes on the SF-12v2. In semi-structured interviews, they justified their assessments. We quantitatively analyzed associations of vignette assessments with individual characteristics using linear regression models and qualitatively analyzed assessment justifications.

Results Patients' age and disease were significant ($p < 0.05$) predictors for ten and seven vignette assessments, respectively. Older patients assessed vignettes being less extreme; patients diagnosed with MS rated them more positively. Overall, adjusted R^2 values ranged from 0.033 to 0.172. Qualitatively, most of the ratings were based on the evaluation of symptoms or impairments in daily life. Fewer participants ranked different dimensions of HRQoL in a personal hierarchy or assumed impairments beyond the vignette description.

Conclusions The understanding of HRQoL may vary substantially and is associated with individual characteristics, individual assessment strategies, and probably other intrinsic factors as explained variance was quite low. Therefore, usage of generic instruments only allows for limited comparison across groups.

Keywords Health-related quality of life · Anchoring vignettes · Mixed-methods · Chronic diseases · Multiple sclerosis · Psoriasis

---

✉ Valerie Andrees
v.andrees@uke.de

[1] Institute for Health Services Research in Dermatology and Nursing (IVDP), University Medical Center Hamburg-Eppendorf (UKE), Martinistraße 52, 20246 Hamburg, Germany

[2] Department of Health Sciences, Competence Center Health, Hamburg University of Applied Sciences (HAW), Lohbrügger Kirchstrasse 65, 21033 Hamburg, Germany

[3] Institute of Neuroimmunology and Multiple Sclerosis (INIMS), University Medical Center Hamburg-Eppendorf (UKE), Martinistraße 52, 20246 Hamburg, Germany

[4] Department of Neurology, University Medical Center Hamburg-Eppendorf (UKE), Martinistraße 52, 20246 Hamburg, Germany

Background/aim

Health-related quality of life (HRQoL) is an essential outcome for determining treatment success from the patient's perspective. It is of special interest in the field of chronic incurable diseases, where the improvement and preservation of HRQoL are primary treatment goals [1]. Consequently, it is indispensable to measure HRQoL for clinical decision making based on reliable and valid instruments [2–4]. To date, a plethora of validated HRQoL measures, among them the SF-36 and the SF-12 questionnaire, exist. Nonetheless, HRQoL measurement faces some obstacles in practice [2].

These obstacles result from the complexity and dynamics of the target construct [4]. HRQoL has diverse definitions that can be understood differently by diverse groups of individuals, it is closely related to other constructs (e.g., self-assessed state of health) and cannot always be clearly differentiated from them [2, 5, 6]. Especially, the distinction from self-assessed state of health is challenging. Some studies do not distinguish between these constructs [7], which shows the ambiguity of HRQoL. The assessment of such a complex subjective construct is prone to biases caused by different understandings of the construct and the questionnaire [5]. HRQoL assessments may differ from one person to another depending on how they interpret the measured construct as well as single questions and response options in a questionnaire, even though objectively they have a similar state of health [8, 9].

Several personal factors may play a role in the individual understanding of HRQoL, the questionnaires, and their response scales. These include cognitive abilities, sociodemographic factors, and health determinants, especially the presence of a chronic condition. The resulting systematic differences in assessments are described as reporting heterogeneity or differential item functioning (DIF) [10, 11]. DIF can cause difficulties in cross-country comparisons of self-reported HRQoL. Yet comparisons across different groups within one country might also be biased by such factors [12], as individuals within one country can be very heterogeneous [13]. In particular, the presence of a chronic disease can alter the reference frame and assessment process of HRQoL and complicates comparison between individuals with different chronic diseases [11]. Various studies examined the influence of this phenomenon on the self-assessment of health states [7, 14, 15], but there is little research on the self-assessment of HRQoL. Therefore, the present study aims to contribute to a better understanding of the individual reference frame and assessment process of HRQoL.

To better understand reporting behavior in self-assessment of health and HRQoL, anchoring vignettes are a recommended method [9, 16, 17]. Vignettes are short descriptions of hypothetical individuals relating to the construct of interest [18]. The vignettes used in this study describe fictive persons with more or less severe impairments in a particular HRQoL domain of the SF-12v2 questionnaire. Using this approach, patients rate their own status and additionally, the status of the vignettes on the same scale. The latter function as reference values, making the measurement more objective [11]. This approach supports the understanding and interpretation of the individual assessment process of HRQoL by adding another individual reference frame.

This study investigates (1) quantitatively how patients with a chronic disease assess vignettes in terms of HRQoL on the SF-12v2 questionnaire and (2) qualitatively their reasoning for the assessment of vignettes.

Methods

Study design, participants, and setting

We carried out an exploratory cross-sectional study based on a mixed-method design. Participant inclusion criteria were being diagnosed with a chronic condition, either psoriasis or multiple sclerosis (MS), and being at least 18 years of age. Severe cognitive impairments were grounds for exclusion. The recruitment of patients took place at two disease-specific outpatient centers, both located at the University Medical Center Hamburg-Eppendorf (UKE), Germany. Inclusion criteria were checked by medical doctors and interested patients were referred to a scientist on site. Given the exploratory character of this study, we aimed to recruit 100 patients, 50 diagnosed with MS and 50 diagnosed with psoriasis.

Development of the method

Vignettes were developed in a multistage procedure, informed by international recommendations and previous research [6, 10, 16, 18]. Semi-structured interviews with nine patients were conducted to determine typical HRQoL impairments. These impairments were utilized to construct the vignettes to ensure that participants could empathize with the examples. This was important for *response consistency*, i.e., that participants use the same rating standard for self-assessment and vignette assessment [10]. Pre-tests with a convenience sample of seven healthy individuals optimized the wording of the vignettes in order to maximize *vignette equivalence*, i.e., that all participants understand the descriptions in the same way [10]. The vignettes were specifically developed for the German SF-12v2, two vignettes for each domain of the questionnaire. In each domain, one vignette

represented being affected more severely than the other. This was achieved by using substantially different concrete depictions of duration and severity of symptoms, i.e., number of days and extent of impairment in daily life. The content of vignettes corresponded to the respective questions of the SF-12v2, i.e., vignettes included concrete information on frequency and duration of impairment when this was asked for in the respective question. To facilitate empathy, the vignettes were worded to describe women for female participants and conversely for male participants. Examples of vignettes, translated into English, are presented online in Appendix I.

Data collection

Data was collected between July and December 2017. We conducted semi-structured, guideline-based interviews including answering of questionnaires. The interviews were structured as follows: First, participants filled in the SF-12v2 questionnaire to assess their own HRQoL. Second, participants assessed HRQoL of vignettes on the SF-12v2. Participants were instructed to associate the examples with their own disease and situation. To reduce bias, the order in which participants received the less or more affected vignette per domain was randomized. For each vignette assessment we used the think-aloud method [19] to gain insights into the underlying decision process of the participant. Third, participants were asked to describe their overall assessment strategies. Finally, a sociodemographic questionnaire was completed. There was no missing data in the questionnaires, because data was collected in face-to-face interviews. All interviews were audio-recorded.

Measurements

SF-12v2

The SF-12v2 is a shorter form of the SF-36 questionnaire in its second version. The validated instrument is available in many languages and measures HRQoL from the patients' perspective. It consists of 12 items covering the following eight domains: general health (GH), physical functioning (PF), social functioning (SF), role-physical (RP), role-emotional (RE), mental health (MH), bodily pain (BP) and vitality (VT). The raw scores of each domain range from 0-100, where higher scores indicate better HRQoL. Domain scores can be summarized into a physical component summary (PCS) and a mental component summary (MCS) [20, 21].

Sociodemographic questions

For the assessment of sociodemographic characteristics, the following variables were collected in written form: gender,

year of birth, family status, level of education, housing situation, year of diagnosis, diagnosed depression, and the presence of other secondary diagnoses.

Statistical analysis

Sociodemographic characteristics and vignette ratings per item and domain were summarized using descriptive statistics. PCS and MCS were calculated for the participants' self-reported HRQoL according to the manual by QualityMetric [22] based on 1998 normative data of the U.S. general population. To analyze the association between sociodemographic characteristics and vignette assessments, stepwise backward linear regression models, with probability to enter $p = 0.05$ and probability to remove $p = 0.10$, were conducted. As the vignettes describe fictive examples of different individuals in each domain, no component scores could be derived for the vignettes, and vignette assessments were analyzed separately on domain level. In addition, we were interested in the difference of the vignette assessments for each domain and determined the assessment difference (*"score vignette 1—score vignette 2"*) as further outcome variables. Therefore, 24 separate regression models were conducted. All independent variables including sociodemographic factors, MCS and PCS (4 continuous, 8 dichotomous) remained the same in every regression model (see Fig. 1). The order of being presented with vignettes (vignette describing a less or more severely affected status first) was considered as a confounder. Assumptions of linear regression analyses were tested in advance. Homoscedasticity and normal distribution of residuals were checked graphically. The latter was additionally verified using Kolmogorov–Smirnov tests. The absence of multicollinearity was verified by VIF and tolerance. For the interpretation of results, Bonferroni correction of p-values was applied to account for inflation of nominal α-level for multiple significance. Therefore, the significance levels of the regression models were adjusted to the number of tests. Statistical analyses were performed with IBM SPSS Statistics V23 (IBM Inc, Armonk, NY, USA).

Qualitative analysis

Qualitative analysis examined patients' justifications of their vignette rating. Patients' justification strategies allow us to make inferences about patients' individual reference frame and assessment process of HRQoL and about aspects and dimensions of relevance. For this analysis an explanatory sequential design was chosen [23], i.e., qualitative analysis was performed based on the results of the quantitative analyses. For practical reasons, we focused on the four vignette assessment regression models with highest explained variance. The case sampling for the qualitative analysis was purposeful [24]. Concentrating on the quantitatively most
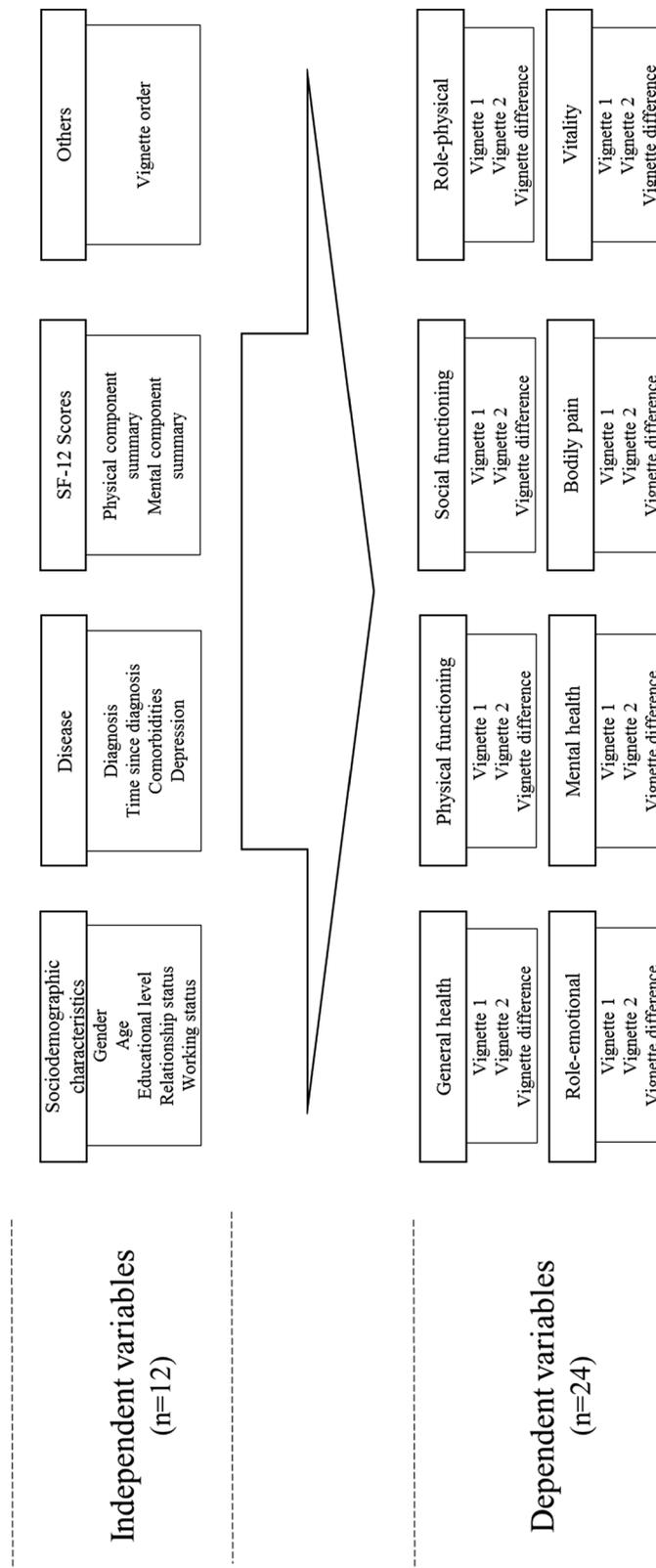
Fig. 1 Independent variables and dependent variables included in regression models, sorted into categories

highly associated variable, extreme or deviant case sampling was performed, selecting participants at the extremes of this variable and covering a broad spectrum of individuals in between [25]. Interviews of selected cases were transcribed verbatim.

The reasoning of vignette ratings was examined to ascertain why the selected participants rated the same vignettes differently. We explored why patients chose a particular response option and not an adjacent one. For the qualitative data analysis, an integrated approach was chosen, combining inductive and deductive coding [26]. With this method, a theoretical framework was built deductively beforehand and was then complemented with inductive coding. This procedure enhances the objectivity of the study and helps to develop a theory about the object of investigation [24]. Assessment justifications and subordinated themes were then generated inductively. In addition, the overall assessment strategy was analyzed separately. The analysis started with 12 interviews and was then gradually added to until theoretical saturation in the overall reasoning for vignette assessment was achieved [27]. For qualitative analysis, MAXQDA Analytics Pro (Verbi GmbH, Berlin, Germany) was used.

Ethics

The study was carried out in accordance with the Code of Ethics of the Declaration of Helsinki and was approved by the ethics committee of the Hamburg Medical Chamber (reference number PV5561). Participants provided written informed consent before participation. No intervention was undertaken.

Results

Study participants

In total, 100 patients participated; 50 diagnosed with MS and 50 diagnosed with psoriasis. Age ranged from 21 to 83 years, with a mean of 45.73 ($\pm$14.63) years. Fifty-two participants were female. Educational level was evenly distributed, with 48 participants with a higher level (at least 12 school years) and 52 with a lower level of education. HRQoL (PCS and MCS) was lower than in the general population [22]. Mean PCS was 43.90 ($\pm$11.07), with MS patients scoring slightly lower (42.68 ± 10.7) than psoriasis patients (45.13 ± 11.33). MCS had a mean of 45.63 ($\pm$12.39), with MS patients (45.48 ± 11.68) and psoriasis patients (45.77 ± 13.18) scoring similarly (Table 1).

Vignette assessment

For each domain of the SF-12, participants rated two different vignettes. These assessments were transformed into scores, ranging from 0–100. As intended, Vignette 1 of each domain, describing a less severely affected patient, was rated with higher mean values of HRQoL than Vignette 2 describing a more severely affected. Mean scores of Vignettes 1 ranged from 51.13 ($\pm$21.03) for RP to 72.25 ($\pm$19.30) for BP. Mean scores of Vignette 2, describing more severely affected patients, ranged from 21.13 ($\pm$15.04) for RP to 43.75 for GH ($\pm$20.75). Mean difference of the two vignette assessments of the respective domain was the lowest for the MH domain (25.00 ± 19.05) and the highest for the BP domain (49.00 ± 25.85) (Table 2). The frequencies of chosen response categories for the vignette ratings are online in Appendix II.

Regression analyses

Testing the assumptions for linear regression revealed satisfactory results. Stepwise backward regression analyses showed that ten of the 16 vignette assessments were significantly associated with participant's age. Being older negatively influenced the assessment of seven vignettes describing, in the main, less severely affected patients and positively influenced the assessment of three vignettes describing more severely affected patients. The participant's disease predicted seven vignette assessments. In comparison to a diagnosis of psoriasis, a diagnosis of MS altered vignette assessments towards more positive ratings. Being diagnosed with depression as secondary diagnosis, in turn, led to more negative vignette assessments. After adjusting significance thresholds according to Bonferroni, age remained significant in three models, disease, and MCS in one model each (Table 3).

Regression models explaining the difference between vignette assessments per domain had less explanatory variance than models explaining single vignette assessments. Yet, age still influenced the assessment differences in six domains. Being older lowered the differences between Vignettes 1 and 2 in those six domains, which means that with the increasing age the less severely affected examples in Vignettes 1 were rated as worse, and the more severely affected cases in Vignettes 2 were rated as better. Further details can be seen in Table 4.

Qualitative analysis

Based on the results of the regression analyses, the four vignette assessments with the highest explained variance were chosen for additional qualitative analysis: GH Vignette 1 (adj. $R^2 = 0.172$), PF Vignette 2 (adj. $R^2 = 0.167$),

Table 1 Descriptive statistics of the study sample ($N=100$)

Variable	N (%)	Mean (SD)
Disease		
MS	50	–
Psoriasis	50	–
Gender		
Female	52	–
Male	48	–
Educational level		
High	48	–
Low	52	–
In a relationship		
Yes	40	–
No	60	–
Working		
Yes	67	–
No	33	–
Comorbidities		
Yes	63	–
No	37	–
Diagnosed with depression		
Yes	12	–
No	88	–
SF-12 scores		
PCS	–	43.90 (11.07)
MCS	–	45.63 (12.39)
Age (years)	–	45.73 (14.63)
Time since diagnosis (years)	–	15.25 (13.90)

Subsample for qualitative analysis ($N=17$)		
Variable	N (%)	
Disease		
MS	7 (41.2)	
Psoriasis	10 (58.8)	
Gender		
Female	8 (47.1)	
Male	9 (52.9)	
Age (years)	–	49.77 (18.77)

N number of patients; *SD* standard deviation; *PCS* physical component summary

MCS mental component summary; *MS* multiple sclerosis

RE Vignette 1 (adj. $R^2 = 0.141$), and BP Vignette 1 (adj. $R^2 = 0.148$). Translations of the four selected vignettes can be found online in Appendix I. In addition to the four vignette assessments, the overall assessment strategy was analyzed.

As participants' age was the strongest predictor of vignette assessment, cases for qualitative analysis were selected based on patients' age. We included extreme cases (21 years, 83 years), complemented with 15 cases that were widely distributed in between. Mean age was 49.76 years

(± 18.77). Of the 17 participants included in the qualitative analysis, eight were female and seven were diagnosed with MS (Table 1).

The analysis of patients' descriptions of their overall assessment strategy revealed that three (sometimes combined) overall assessment patterns were used to assess the vignettes on the scales of the questionnaire. The most common strategy, used by all 17 participants, was to compare the examples with their own situation. Most participants said they tried to imagine how they would feel in the same

Table 2 Results of vignette assessments per domain

Domain	Vignette 1				Vignette 2				Vignette difference			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
General health	0.00	100.00	62.10	25.76	0.00	100.00	43.75	20.75	− 40.00	100.00	27.35	27.72
Physical functioning	0.00	100.00	69.75	23.91	0.00	75.00	21.75	21.22	− 50.00	100.00	48.00	26.99
Social functioning	0.00	87.50	51.13	21.03	0.00	62.50	21.13	15.04	− 25.00	75.00	30.00	21.02
Role-physical	0.00	87.50	54.63	22.66	0.00	87.50	22.75	19.25	− 62.50	87.50	32.88	30.59
Role-emotional	0.00	100.00	72.25	19.30	0.00	75.00	26.25	22.30	− 75.00	100.00	49.00	25.85
Mental health	12.50	75.00	55.63	14.69	0.00	62.50	30.63	14.03	− 50.00	75.00	25.00	19.05
Bodily pain	0.00	100.00	69.75	18.90	0.00	100.00	21.75	16.53	− 50.00	100.00	48.00	25.54
Vitality	0.00	100.00	64.75	18.50	0.00	100.00	25.25	18.63	− 100.00	100.00	39.50	27.33

Min minimum; *Max* maximum; *SD* standard deviation

situation and to what extent the situation would impair them. Eight stated they could relate to the described examples because they had experienced similar situations. Two participants said relating to the situation was difficult because they would cope differently than the patient described in the example, but still tried it. In addition, most participants described using a logical reasoning process in assessing the vignettes: first reading questions and examples, then selecting the most important information. Some participants mentioned the time frame, e.g., the time span of the impairment, as the key factor, while others focused on the type and level of impairment. Several participants stated that they compared the two example patients in each domain with each other to rate which of the two was more impaired.

For single vignette assessments, we analyzed the verbalized decision process and explored why participants decided for a particular response option and not an adjacent one. We found that the selection between adjacent response categories (e.g., "some of the time" vs. "a little of the time" or "good" and "very good") was not justified in a way that allowed for distinction between them. Therefore, we explored why some participants rated a vignette as positive while others rated it as negative. In the following, we display the most frequently used response categories to the respective justification in brackets. The justifications for single vignette ratings were assigned to four different strategies: (1) evaluation of symptom quality and severity, (2) evaluation of impairments in daily life, (3) ranking different dimensions of HRQoL in a personal hierarchy, and (4) assuming impairments beyond the described vignettes (Table 5). The most commonly applied strategies were the evaluation of symptoms and impairments in daily life. The strategies were not associated with sociodemographic factors.

Concerning the less severely affected *GH Vignette 1*, rather negative assessment (mostly "fair") was most often justified through the evaluation of the symptoms, for example being sad or thoughtful. Some participants assumed the presence of impairments beyond the mere description,

such as a suspected long-term condition. In contrast, rather positive assessments (mostly "very good" and "good") were most often justified via the nonexistence or short duration of impairments in daily life. In addition, some participants ranked mental health problems as less important than physical health problems.

In the assessment of the more severely affected *PF Vignette 2*, only the evaluation of symptoms and the evaluation of impairments in daily life figured in justifications. Focusing on the described symptoms led to a rather negative rating (mostly "Yes, limited a lot") of the vignette. Patients mainly considered the type of pain (e.g., knee pain) and the duration of pain in their negative assessment. The impairments in daily life were evaluated both negatively and positively. A negative rating was justified by citing the need for support, the use of the elevator, and the abandonment of a hobby. A more positive rating (mostly "Yes, limited a little") was justified by stating that the impairment could be worse, e.g., needing a wheelchair. In addition, some participants rated receiving support from family and friends as positive and as not an impairment.

For the less severely affected *RE Vignette 1*, a rather positive rating (mostly "a little of the time" and "some of the time") was most often justified via the evaluation of impairments in daily life. Participants focused on the short time span, the ability to work and the support of family and friends. One patient rated support negatively. In a negative assessment, the evaluation of symptoms was a driving factor. Especially, psychological stress, sadness, and lethargy were mentioned as justifications for a negative rating (mostly "all of the time" and "most of the time"). One participant assumed the existence of a justification beyond the textual description of the vignette by interpreting a joylessness at work and argued that work should be fun.

Positive ratings (mostly "Not at all" or "A little bit") of *BP Vignette 1* describing a less severely affected status were either justified via the evaluation of the symptoms or the evaluation of the impairments in daily life. Participants said

Table 3 Standardized regression coefficients in stepwise linear regression: less severely affected Vignette 1 and more severely affected Vignette 2 per domain (N = 100)

Vignette	Age	Disease	Depression	Time since diagnosis	Vignette order	MCS	Working	Educational level	Comorbidities	Overall model summary		
Ref. cat.	(Years)	Psoriasis	No	(Years)	Less severe first	(Score)	No	Low	No	R^2	Adj. R^2	SE
General health 1	−0.254**	0.270**	–	–	0.224*	–	–	−0.192	–	0.206	0.172	23.44
General health 2	−0.231*	0.217*	−0.228*	–	–	0.184	–	−0.191	0.184	0.212	0.161	19.00
Physical functioning 1	−0.218*	0.275**	–	–	0.170	–	–	–	–	0.150	0.124	22.39
Physical functioning 2	0.287***†	0.355***†	–	–	–	–	–	–	–	0.183	0.167	19.37
Social functioning 1	−0.199	0.176	–	–	0.216*	–	–	0.196	–	0.145	0.109	17.45
Social functioning 2	0.260**	–	−0.215*	–	−0.165	–	–	–	–	0.115	0.087	17.80
Role-physical 1	−0.245*	0.225*	–	0.245*	–	–	–	–	−0.191	0.127	0.090	20.06
Role-physical 2	0.206*	–	–	–	–	–	–	–	–	0.042	0.033	14.80
Role-emotional 1	−0.390***†	0.293**	–	0.271*	–	–	–	–	**–**	0.176	0.141	21.01
Role-emotional 2	–	–	–	–	–	–	–	–	–	–	–	–
Mental health 1	–	0.168	–	–	–	0.222*	−0.241*	–	–	0.140	0.113	13.84
Mental health 2	–	–	−0.287***	–	–	–	**–**	–	–	0.082	0.073	13.51
Bodily pain 1	−0.490***†	–	–	0.234*	–	–	–	–	–	0.165	0.148	17.81
Bodily pain 2	–	0.195*	–	–	–	0.296***†	–	–	–	0.124	0.106	21.08
Vitality 1	−0.239*	–	−0.258**	–	–	–	–	–	–	0.127	0.109	17.84
Vitality 2	–	–	–	–	–	–	–	–	–	–	–	–

MCS mental component summary; R^2 coefficient of determination; *adj.* adjusted; *SE* standard error

bold beta values: statistical significant; *p* value: < 0.05, **p-value: < 0.01; † Remains significant at nominal p-value ≤ 0.05 after Bonferroni correction for 16 significance tests

Nonincluded predictors are denoted by "–"; predictors not included in any model are not displayed (gender, relationship status, physical component summary)

Table 4 Standardized regression coefficients in stepwise linear regression predicting the vignette difference per domain (N = 100)

Vignette difference	Age	Disease	Vignette order	Depression	Time since diagnosis	MCS	Working	Gender	Overall model summary		
Ref. cat.	(Years)	Psoriasis	Less severe first	No	[Years]	[Score]	No	Female	R^2	Adj. R^2	SE
General health	–	–	–	0.275*†	–	–	–	–	0.076	0.066	26.78
Physical functioning	– 0.401**†	–	–	–	–	–	–	–	0.161	0.152	42.85
Social functioning	– 0.355**†	–	– 0.192*	–	–	–	–	–	0.151	0.134	25.44
Role-physical	– 0.415**†	0.192	0.189	–	0.204	–	–	–	0.163	0.128	19.63
Role-emotional	– 0.291**†	0.210*	–	–	–	0.188*	–	–	0.167	0.141	28.34
Mental health	–	0.174	–	–	–	–	0.256*	0.218*	0.110	0.082	18.25
Bodily pain	– 0.329**†	–	–	–	0.302*	– 0.227*	–	–	0.126	0.099	24.54
Vitality	– 0.262**	–	–	–	–	–	–	–	0.069	0.059	24.78

MCS mental component summary; R^2 coefficient of determination; *adj.* adjusted; *SE*: standard error

bold beta values: statistical significant

*p-value: < .05, **p-value: < .01; † Remains significant at nominal p-value ≤ .05 after Bonferroni correction for 8 significance tests; nonincluded predictors are denoted by "-"; predictors not included in any model are not displayed (relationship status, physical component summary, comorbidities, working)

that pain was short-term, not severe, and painkiller treatment was successful, consequently working was possible. Negative rating (mostly "moderate") was rarer and justified by either inferring the existence of impairments beyond the vignettes, such as psychological stress, or by ranking pain as a very important component of HRQoL.

Discussion

This study has shown that vignette assessments on the SF-12 were related to sociodemographic factors and to individual reasoning strategies during the assessment process. The results provide insights into the HRQoL construct but the understanding of the construct is still limited and further research is needed.

Individual characteristics predicting vignette assessments

The quantitative analysis indicates that different individual factors were related to the assessment of the same vignettes. Especially, participants' age- and disease-specific factors had an impact on individual assessments. Contrary to findings of other studies, related constructs such as self-rated health or pain assessment, participants' gender [7, 28, 29] and educational level [15, 29] did not influence assessments.

In detail, we found that younger patients rated vignettes describing more severely affected patients more negatively and vignettes describing less severely affected patients more positively. In other words, the difference between vignette ratings in each domain decreased with age. These results dovetail with a previous research finding that older people

show a higher steadiness of emotional experiences, while younger people are more reactive to unforeseen events in daily life when assessing their own well-being [30, 31]. Our results suggest that this might not only be in self-assessment but also in assessment of external references such as vignettes which implies that individual understandings of the construct differ.

Participants diagnosed with MS had slightly worse self-rated HRQoL, which is consistent with the slightly higher disability weights in the global burden of disease study for patients with MS compared to psoriasis [32]. While reporting worse HRQoL for themselves, MS patients rated the HRQoL of the patients described in the vignettes as better on average than psoriasis patients. This may be because they take their own HRQoL as a reference value, thus contrasting their own and HRQoL states in the vignettes. The differences between disease groups suggest that different reference frames are used by those groups, leading to difficulties in comparing their assessments. Disease and age remained significant predictors in some models after Bonferroni adjustment. Notably, age remained significant in all models investigating vignette difference. Yet, it needs to be considered that statistical power of all models decreased due to Bonferroni adjustment. Given the exploratory character of this study, findings need to be verified in confirmatory studies. Accordingly, results can serve as a basis for deriving hypotheses on patients' individual reference frame of HRQoL.

Reasoning for vignette assessments

To be able to integrate and connect quantitative and qualitative analyses, the qualitative sampling was based on the

Table 5 The four reasoning strategies for evaluating HRQoL in vignettes with exemplary quotes

(1) Evaluation of symptom quality and severity	Bodily pain vignette 1 Role-emotional vignette 1	Female, 83 years, MS Male, 38 years, psoriasis	"When she took the painkiller, she was fine. Additionally, it was only two days. A little bit, yes."(JT03) "That one is clearly worse off. Well, he is often sad. Actually, mainly…" (JT57)
(2) Evaluation of impairments in daily life	Physical functioning vignette 2 General health vignette 1	Female, 27 years, psoriasis Female, 64 years, MS	"Well, she can do small tasks, but as she had to give up her hobbies and is impaired in window cleaning and some other things, I would tend to limited a lot." (JT29) "Good … although she has MS. She is employed and everything works out in daily life. And cancelling a visit to the cinema is not that bad in my opinion."(VA42)
(3) Ranking different dimensions of HRQoL in a personal hierarchy	General health vignette 1 General health vignette 1	Male, 52 years, MS Female, 45 years, MS	"Mr. Schulz is very well in my opinion, because he can do everything he wants. Only, because his thoughts drag him down somehow… But actually he is well off."(VA18) "Well, would say very good, actually. Because there is nothing written about impairments. Except for this… [psychological problem]"(VA51)
(4) Assuming impairments beyond the described vignettes	Bodily pain vignette 1 General health vignette 1	Male, 64 years, psoriasis Female, 21 years, MS	"…and then he took painkillers and went to work. Afterwards he was without pain. That means he has a problem at work. A psychological problem. And this only if certain things occur. In other words: stress." (JT09) "… Mrs. Schulz is worrying about her disease, which makes her feel down, because she feels excluded and compares herself with her friends…"(VA29)

MS multiple sclerosis

quantitative results. As age and disease had the highest predictive value, we assumed that the justification of assessment strategies would differ between young and old and between MS and psoriasis patients. However, we could not detect these predictors in qualitative reasoning giving expression to the assessments, but instead found justification strategies to differ between domains.

Depending on the SF-12 domain and on whether the rating of the single vignettes was overall positive or negative, four different strategies were used to assess the vignettes. However, these strategies were not related to whether the rating was extreme or moderate. So, these strategies did not explain why participants chose one response option rather than an adjacent one (e.g., "good", "very good", "excellent"), as adjacent response options were justified in a very similar or even identical manner. However, quantitative analysis showed that in cases where younger participants rated "good" overall health, older ones rated "excellent", even though the qualitative assessment strategy did not differ. This leads one to conclude that not only are assessments affected by the combination of sociodemographic factors and justification strategies, but also by additional factors that predict why participants chose one answer rather than an adjacent one. In this context, the stableness of personal emotional experience [30, 31] might have influenced the assessment and could explain assessment differences with age.

The analyses demonstrated that justification patterns were not uniform but changed with HRQoL domains of the SF-12 and each domain exhibited different patterns. Individual participants also changed strategy between domains. Therefore, the interaction between sociodemographic characteristics and justification strategies becomes even more complex with different HRQoL domains.

This study shows that the individual assessment process of patients with chronic conditions is complex. On the one hand, it is influenced by the participants' sociodemographic and health-related characteristics, especially age and disease. On the other hand, different domains of HRQoL are evaluated by recourse to different assessment strategies, the use of which does not depend on sociodemographic characteristics. We assume that additional determinants influence the assessment process of HRQoL. These determinants could be personality traits, like openness and emotional coping style, which have been found to predict HRQoL, too [33].

Strengths and limitations

The major strength of this study is its mixed-method design. The combination of sociodemographic aspects and personal assessment strategies provided a broader picture and improved understanding of the individual assessment process of HRQoL. In addition, face-to-face interviews provided a dataset without missing values.

As the tasks for participants were rather complex, patients with severe cognitive impairments were excluded. As these patients may also have lower HRQoL, the difference between disease groups in our study may have been smaller than that in a representative sample.

Due to ethical considerations, no data on income were collected, although other studies have found that income was a key factor behind reporting heterogeneity [8, 29], and therefore, for different individual reference frames of HRQoL. Further research on this topic should include income as an independent variable.

A further strength of this study is its extensive and comprehensive development of the vignettes for the SF-12. Vignette equivalence, i.e., the assumption that all participants understand the descriptions therein in the same way, was recognized in qualitative analysis and was only violated when patients inferred information beyond the descriptive text, which was rare. Response consistency, i.e., using the same rating standard for self-assessment and vignette assessment, could be assumed for nearly every patient, as participants were told to associate the examples with themselves and confirmed this when reporting on their assessment strategies.

Conclusion

This exploratory research revealed that individual understanding and assessment processes of HRQoL may substantially vary and is predicted by sociodemographic factors, individual assessment strategies, and probably other intrinsic factors. It provides a useful basis for further research on those predictors as it shows that the assessment of vignettes is a reasonable approach for improving our understanding of HRQoL.

Generic instruments aim to overcome the limitations of disease-specific instruments and make comparisons across different diseases possible. However, this study shows that generic instruments, too, only allow for limited comparison, because the HRQoL construct is understood differently by different disease groups.

References

1. Greenfield, J., Hudson, M., Vinet, E., Fortin, P. R., Bykerk, V., Pineau, C. A., et al. (2017). A comparison of health-related quality of life (HRQoL) across four systemic autoimmune rheumatic diseases (SARDs). *PLoS ONE, 12*(12), e0189840. https://doi.org/10.1371/journal.pone.0189840.

2. Lin, X.-J., Lin, I.-M., & Fan, S.-Y. (2013). Methodological issues in measuring health-related quality of life. *Tzu Chi Medical Journal, 25*(1), 8–12. https://doi.org/10.1016/j.tcmj.2012.09.002.

3. Badia, X., & Herdman, M. (2001). The importance of health-related quality-of-life data in determining the value of drug therapy. *Clinical Therapeutics, 01*(23), 168–175. https://doi.org/10.1016/S0149-2918(01)80039-4.

4. Ahmed, S., Berzon, R. A., Revicki, D. A., Lenderking, W. R., Moinpour, C. M., Basch, E., et al. (2012). The use of patient-reported outcomes (PRO) within comparative effectiveness research: Implications for clinical practice and health care policy. *Medical Care, 12*(50), 1060–1070. https://doi.org/10.1097/MLR.0b013e318268aaff.

5. Sosnowski, R., Kulpa, M., Zietalewicz, U., Wolski, J. K., Nowakowski, R., Bakula, R., et al. (2017). Basic issues concerning health-related quality of life. *Central European Journal of Urology, 70*(2), 206–211. https://doi.org/10.5173/ceju.2017.923.

6. Crane, M., Rissel, C., Greaves, S., & Gebel, K. (2016). Correcting bias in self-rated quality of life: an application of anchoring vignettes and ordinal regression models to better understand QoL differences across commuting modes. *Quality of Life Research, 25*(2), 257–266. https://doi.org/10.1007/s11136-015-1090-8.

7. Schneider, U., Pfarr, C., Schneider, B. S., & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. *The European Journal of Health Economics, 13*(3), 251–265. https://doi.org/10.1007/s10198-011-0301-7.

8. Etilé, F., & Milcent, C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from France. *Health Economics, 15*(9), 965–981. https://doi.org/10.1002/hec.1164.

9. Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics, 17*(3), 351–375. https://doi.org/10.1002/hec.1269.

10. King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(01), 191–207. https://doi.org/10.1017/S000305540400108X.

11. Knott, R. J., Lorgelly, P. K., Black, N., & Hollingsworth, B. (2017). Differential item functioning in quality of life measurement: An analysis using anchoring vignettes. *Social Science and Medicine, 190,* 247–255. https://doi.org/10.1016/j.socscimed.2017.08.033.

12. Abdin, E., Subramaniam, M., Picco, L., Pang, S., Vaingankar, J. A., Shahwan, S., et al. (2017). The importance of considering differential item functioning in investigating the impact of chronic conditions on health-related quality of life in a multi-ethnic Asian population. *Quality of Life Research, 26*(4), 823–834. https://doi.org/10.1007/s11136-016-1418-z.

13. Buchcik J, Westenhöfer J, Fleming M, Martin CR (2017) Health-related quality of life (HRQoL) among elderly Turkish and polish migrants and German Natives: The role of age, gender, income, discrimination and social support. Discrimination and social support, people's movements in the 21st century—risks, challenges and benefits, Dr. Ingrid Muenstermann (Ed.). *InTech*. https://doi.org/10.5772/66931.

14. O'Doherty, M. G., French, D., Steptoe, A., & Kee, F. (2017). Social capital, deprivation and self-rated health: Does reporting heterogeneity play a role? Results from the English Longitudinal Study of Ageing. *Social Science and Medicine, 179,* 191–200. https://doi.org/10.1016/j.socscimed.2017.03.006.

15. Hanandita, W., & Tampubolon, G. (2016). Does reporting behaviour bias the measurement of social inequalities in self-rated health in Indonesia? An anchoring vignette analysis. *Quality of Life Research, 25*(5), 1137–1149. https://doi.org/10.1007/s11136-015-1152-y.

16. Au, N., & Lorgelly, P. K. (2014). Anchoring vignettes for health comparisons: an analysis of response consistency. *Quality of Life Research, 23*(6), 1721–1731. https://doi.org/10.1007/s11136-013-0615-2.

17. Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2014). Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases. *Oxford Bulletin of Economics and Statistics, 76*(5), 643–666. https://doi.org/10.1111/obes.12039.

18. Grol-Prokopczyk, H., Freese, J., & Hauser, R. M. (2011). Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior, 52*(2), 246–261. https://doi.org/10.1177/0022146510396713.

19. Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*. https://doi.org/10.26522/brocked.v12i2.38.

20. Jenkinson, C., Layte, R., Jenkinson, D., Lawrence, K., Petersen, S., Paice, C., et al. (1997). A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *Journal of Public Health Medicine, 19*(2), 179–186.

21. Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., et al. (1998). Cross-validation of Item selection and scoring for the SF-12 health survey in nine countries: Results from the IQOLA project. *Journal Clinical Epidemiology, 51*(11), 1171–1178.

22. Mariush, M. E. (2012). *User's manual for the SF-12v2 health survey* (3rd ed.). Lincoln: QualityMetric Incorporated.

23. Kuckartz, U. (2017). Datenanalyse in der mixed-methods-Forschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69*(S2), 157–183. https://doi.org/10.1007/s11577-017-0456-z.

24. Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet, 358*(9280), 483–488. https://doi.org/10.1016/S0140-6736(01)05627-6.

25. Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77–100. https://doi.org/10.1177/1558689806292430.

26. Bradley, E. H., Curry, L. A., & Devers, K. J. (2007). Qualitative data analysis for health services research: Developing taxonomy, themes, and theory. *Health Services Research, 42*(4), 1758–1772. https://doi.org/10.1111/j.1475-6773.2006.00684.x.

27. Curry, L. A., Nembhard, I. M., & Bradley, E. H. (2009). Qualitative and mixed methods provide unique contributions to outcomes research. *Circulation, 119*(10), 1442–1452. https://doi.org/10.1161/CIRCULATIONAHA.107.742775.

28. Chan, A., Malhotra, C., Do, Y. K., Malhotra, R., & Østbye, T. (2011). Self reported pain severity among multiethnic older Singaporeans: Does adjusting for reporting heterogeneity matter? *European Journal of Pain, 15*(10), 1094–1099. https://doi.org/10.1016/j.ejpain.2011.05.006.

29. Hirve, S., Verdes, E., Lele, P., Juvekar, S., Blomstedt, Y., Tollman, S., et al. (2014). Evaluating reporting heterogeneity in self-rated health Among adults aged 50 years and above in India. *Journal of Aging and Health, 26*(6), 1015–1031. https://doi.org/10.1177/0898264314535634.

30. Röcke, C., Li, S.-C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging, 24*(4), 863–878. https://doi.org/10.1037/a0016276.

31. Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., et al. (2011). Emotional experience improves with age: Evidence based on over 10 years of experience sampling. *Psychology and Aging, 26*(1), 21–33. https://doi.org/10.1037/a0021285.

32. Global Burden of Disease Collaborative Network (2017). Global Burden of Disease Study 2016 (GBD 2016) Disability Weights.

33. Pereira-Morales, A. J., Adan, A., Lopez-Leon, S., & Forero, D. A. (2018). Personality traits and health-related quality of life: The mediator role of coping strategies and psychological distress. *Annals of General Psychiatry, 17,* 25. https://doi.org/10.1186/s12991-018-0196-0.

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