Diabetology International (2019) 10:93–101
https://doi.org/10.1007/s13340-018-0371-1

ORIGINAL ARTICLE

The revalidation of the diabetes treatment-related quality-of-life (DTR-QOL) questionnaire in Japan

Hitoshi Ishii[1] · **Hyunchung Ray Kim**[2] · **Bruce Crawford**[2]

Received: 9 April 2018 / Accepted: 24 July 2018 / Published online: 28 August 2018

Abstract

Objectives The objectives of this study were to develop a conceptual framework for the diabetes treatment-related quality of life (DTR-QOL) and validate the new structure of the DTR-QOL.

Methods This study was conducted in two stages. First, items were collated into similar concepts to develop a new conceptual framework for the DTR-QOL. Next, psychometric analyses were conducted. Analyses included a distribution of responses (and domains) with a focus on floor/ceiling effects, item convergent/discriminant validity, internal consistency reliability (where possible), clinical validity, known-group analysis, and concurrent validity with the Diabetes Treatment Satisfaction Questionnaire and SF-8. Known groups were evaluated using "Glycemic control in the last month", "Concerned with weight gain in the last month", "Current health status", and "Degree of communication with clinician". To evaluate the reliability of each scale, test–retest were assessed.

Results The DTR-QOL items were reconstructed into seven domains based on similar concepts: usual activities, pain/discomfort, anxiety/depression, satisfaction, positive feelings, negative feelings, and feeling troubled. Although there were ceiling effects on most items, when patients reported having a hypoglycemic event in the last month, score distributions found fewer ceiling effects indicating that items are able to respond to clinical changes. Internal consistency reliability was met for all scales except satisfaction ($\alpha = 0.69$) and pain/discomfort ($\alpha = 0.66$). Clinical validity, known groups, and concurrent validity were met. Test–retest met acceptable levels for all scales except positive feelings.

Conclusions The restructuring of the DTR-QOL resulted in an acceptable measure with domains that are easily interpreted and allow physicians to better measure more specific impacts on their patients.

Keywords Diabetes · Quality of life · Treatment · Validation · Japan

Introduction

With a growing number of type-2 diabetes mellitus (T2DM) patients in Japan, and its significant impact on patients' quality of life and general health conditions, better ways to manage the disease and monitor patients' feedback via routine care is imperative more than ever [1–3]. Being a chronic condition, T2DM requires a high level of treatment adherence by patients, which is highly associated with treatment satisfaction and patients' overall perception of quality of life (QOL) [4–8].

The diabetes treatment-related quality-of-life (DTR-QOL) questionnaire is a multi-domain patient-reported outcome instrument developed in Japan, to derive results on patients' QOL regarding diabetes treatment, which aligns with the importance of patient-focused disease management by measuring patients' improvement of QOL and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13340-018-0371-1) contains supplementary material, which is available to authorized users.

✉ Hitoshi Ishii
hit3910@gmail.com

Hyunchung Ray Kim
Ray.kim@syneoshealth.com

Bruce Crawford
bruce.crawford@syneoshealth.com

[1] Department of Diabetology, Nara Medical University Hospital, 840 Shijo-Cho, Kashihara, Nara, Japan

[2] Syneos Health, 5F Urbannet Nihonbashi 2-Chome Building, 2-1-3 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

satisfaction of treatment, as part of routine care [9]. Despite its fit-for-purpose nature for T2DM patients, there have been concerns with how to interpret the results in the initial phase, as the initial domain structure of the DTR-QOL was not unidimensional, but a mixture of concepts due to a reliance solely on statistical modeling. In addition, there was no conceptual framework available in the initial stage. According to Rothman et al, when a framework does not exist, it is recommended to develop one to demonstrate how the items and scales are related to one another upon discussion with the developer to better reflect the content and link to the claim [10]. Similarly, the US Food and Drug Administration (FDA) Guidance for Industry states the importance of clearly addressing construct validity along with content validity when assessing adequacy of a patient-reported outcomes measure (PROM) [11].

For this reason, this study first aimed to reconstruct the questionnaire to modify the existing structure into more interpretable and discrete domains, through a conceptual framework exercise and investigate the content validity via traditional psychometric validation methods to understand the performance of each domain and document its properties [12].

Methods

Data set

This analysis utilized the same data set utilized in the original validation study of the DTR-QOL [9], thereby building upon the earlier analyses with a new domain structure. The study population consisted of adult patients with either type 1 or type 2 diabetes. As this is a reanalysis of the previously collected data, additional ethics submission was not required.

The diabetes therapy-related quality-of-life (DTR-QOL) questionnaire is a diabetes-specific instrument developed and validated to assess the influence of diabetes treatment on QoL. The DTR-QOL consisted of 29 items utilizing a 7-point Likert scale (1: strongly agree–7: strongly disagree). The previous exploratory factor analysis resulted in 4 domains: burden on social activities and daily activities, anxiety and dissatisfaction with treatment, hypoglycemia, and satisfaction with treatment [9]. It has been used to measure diabetes patients' QoL in Japan since its development and preliminary validation [13–15]. A copy of the DTR-QOL has been provided in the Supplementary Materials with a brief English translation (this translation is the authors forward translation only and not an official forward–backward translation needed for use in the English language).

Additional questionnaires used in this study include the Diabetes Treatment Satisfaction Questionnaire (DTSQ) [16, 17] and the 8-item short form health survey (SF-8) [18]. The DTSQ is comprised of 8 items that assess satisfaction with treatment as it relates to diabetes. The SF-8 is a generic QoL questionnaire which is scored to create 2 composite domains: physical component summary (PCS) and mental component summary (MCS) scores. The PCS and MCS are normative based, such that they have a mean of 50 and standard deviation of 10, based on the Japanese population.

Phase I: reconstruction of the DTR-QOL

To update the domain structure of the DTR-QOL, a conceptual framework was developed to better capture all the concepts within the DTR-QOL. The DTR-QOL was examined closely at the item level to develop concise groupings of "like concepts" that were more unidimensional than those of the previous study [9]. A conceptual framework, representing the new structure, presented each item in their respective domain. Through close examination of the concepts being measured in each question, the research team grouped items together initially according to FDA guideline [12, 19], followed by developing a domain name that represents the content of the proposed domain.

Phase II: validation of the DTR-QOL

Following the restructuring of items, a psychometric validation on the new structure was conducted to demonstrate its properties and ability to differentiate subgroups. The validation analyses on the DTR-QOL's new structure evaluated its reliability and validity. Analyses included a distribution of responses (and domains) with a focus on floor/ceiling effects, item convergent/discriminant validity, internal consistency reliability of each domain, clinical validity, known-group analysis, and concurrent validity with other instruments. To evaluate the reliability of each scale, test–retest were assessed. Psychometric validation focused more on the magnitude and direction of relationships rather than on statistical significance.

Floor/ceiling effects evaluated the proportion of patients scoring at the lowest/highest possible score, respectively. When a question (or scale) has a large proportion of patients at the highest score (ceiling effect), it is not able to detect patients improving. In contrast, a large proportion of patients scoring the lowest score (floor effect) is not able to detect worsening. These scores needed to be evaluated in regards to the population under investigation.

Item convergent/discriminant validity was used to assess correlations between each item and the domain scores. Items should correlate at a level of 0.40 or greater with their own scale (convergent validity) and be more correlated with their respective domain than another domain (discriminant validity). Items that cross correlate with several domains were

examined closely to understand the item content and domain content. Item–domain selection was determined primarily based on content rather than correlations to ensure interpretability. As item–domain correlations are almost always highly significant, the magnitude of correlation was examined. When items were correlated with their own domain and another domain, we sought physician's advice.

Internal consistency reliability was assessed with Cronbach's alpha and represented how well the questions fit together as a domain. A threshold of 0.70 or greater was usually sought as an indicator of good fit [20]. A standardized alpha coefficient was used to standardize reporting on a 10-item scale. Items were examined for changes in alpha level with that item removed to identify any possible misfitting items. Only domains with three or more items are able to be assessed for internal consistency reliability.

For groups with known differences (known groups), the domain should also demonstrate differences. Known groups were evaluated for "Glycemic control in the last month", "Concerned with weight gain in the last month", "Current health status", and "Degree of communication with clinician". Significance testing was conducted with ANOVA, although the primary focus was on demonstrating a monotonic positive or negative relationship across response options. Concurrent validity was assessed by evaluation of the correlation between the DTR-QOL domains with the DTSQ and the SF-8.

Test–retest were calculated using intraclass correlation coefficients [Shrout–Fleiss ICC(2,1)]. A threshold of 0.70 or greater is considered as minimum criteria for acceptable test–retest reliability.

Results

New domain structure

The DTR-QOL questionnaire items were reconstructed into 7 domains based on similar concepts of the items: usual activities, pain/discomfort, anxiety/depression, satisfaction, positive feelings, negative feelings, and feeling troubled (Fig. 1). "Usual Activities" contains 9 items, and measures the restrictions that the patient experiences in various aspects of their daily life, including interference with the patient's work and daily activities, time and location to travel, diet, and relationships with others. "Pain/Discomfort" contains 4 items and measures the pain/discomfort that the patient may experience in their current diabetes treatment as well as discomfort from symptoms. "Anxiety/Depression" contains 5 items and measures the worries/concerns that patients have with their current diabetes treatment and condition. "Satisfaction" contains 3 items and is related to how satisfied or dissatisfied patients are with their current diabetes treatment

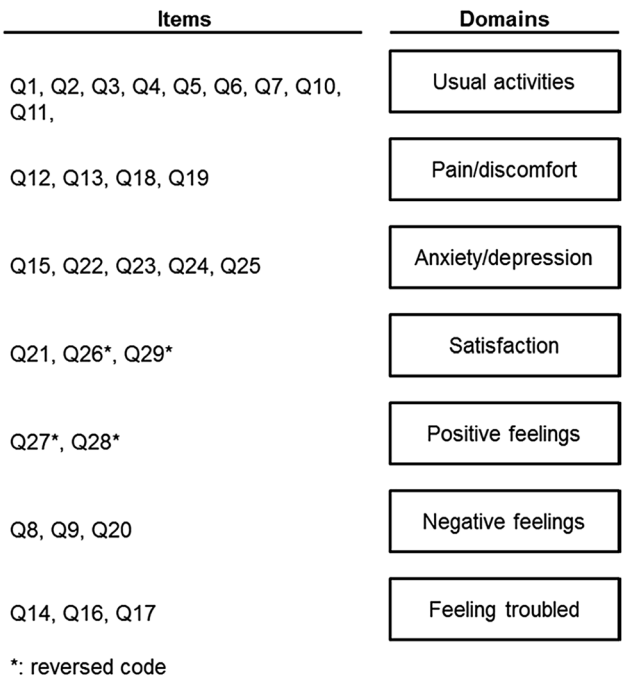

Fig. 1 Reconstructed domains of DTR-QOL

and blood glucose-level control. "Positive Feelings" contain 2 items, measuring the positive thoughts and confidence that patients have for their current and future condition. "Negative Feelings" contain 3 items, measuring the patient's distress on their current diabetes state and impact on their daily life. "Feeling Troubled" measures how troubled the patient feels about their current diabetes state.

Psychometric validation

The DTR-QOL domains were scored on a 0–100 range, with 100 representing the best score, by transforming the summed score to a 0–100 scale. First, items 26–29 were reversely scored to ensure that a high score was considered better. Next, items belonging to the same domain were summed (as shown in Fig. 1), using the 50% rule, a common technique used to impute data in widely used health-related quality-of-life measures. The 50% rule dictates that at least 50% + 1 items must have a valid score in order for the scale to be calculated. Missing values are then replaced by the mean score of available items. The transformed score was developed as (total raw score–lowest possible score)/(range of possible scores).

Item-level distributions were examined at baseline and items experienced a high level of ceiling effects on most items. Item-level distributions were further examined by "*hypoglycemic event in the past month*" (yes/no) to understand the impact of a negative event. The item distributions for the hypoglycemic event group demonstrated a better

distribution of responses, with fewer ceiling effects, demonstrating that the items did change with more severe clinical events.

Item-scale correlations were produced and examined. In the initial correlation matrix produced, it was found that item 11 (*I am constantly concerned about time to manage my current diabetes treatment*) and item 16 (*I am scared because of low blood sugar*) more highly correlated with the usual activities and feeling troubled domains, respectively. After further evaluation of the item content, these items were moved to the new domains. The new item-scale correlation matrix revealed that all items correlated well with their own domain and all items except 2 correlated higher with their own domain than another domain. Although item 15 (*I worry about low blood glucose due to my current diabetes treatment*) correlated more highly with feeling troubled (0.72) than with anxiety/depression (0.63), the content of the item was judged as a better fit with the anxiety/depression domain according to expert opinion. Item 18 (*Symptoms due to low blood sugar are uncomfortable*) also correlated more highly with feeling troubled (0.82) than with pain/discomfort (0.78). It was also determined to leave it in pain/discomfort as that domain better represents the content of the item.

Based on the new domain scaling, internal consistency reliability found most scales to above the common threshold of acceptability (0.70). The pain/discomfort and satisfaction domains were found to be lower than the threshold, at 0.66 and 0.69, respectively (Table 1). Satisfaction was very close to the threshold, and, therefore, was not a major concern. The pain/discomfort domain was lower than desired and should be further examined in another population. As a minimum of three items was required to calculate an alpha coefficient, positive feelings were not able to be assessed for internal consistency reliability.

The examination of scale–scale correlations revealed moderate correlations between all domains, with the highest correlation between pain/discomfort and feeling troubled ($r = 0.76$; Table 2). The correlation was low enough to not represent the measurement of the same concept, yet they were conceptually highly related. This strong relationship may be acceptable as patients may not feel troubled without having pain and/or discomfort from their symptoms and treatment. Therefore, conceptually, one was needed to affect the other.

The DTR-QOL domain summary statistics at baseline and retest are presented in Tables 3 and 4, respectively. Similar to the item statistics, there remained some ceiling effects for pain/discomfort and feeling troubled.

Clinical validity of the DTR-QOL domains was assessed using any reported hypoglycemic event within the last month prior to completion of the questionnaire. All domains reported mean scores in the anticipated direction; patients with a hypoglycemic event were expected to report lower scores on all domains (Table 5). Domains showed significant differences in all domains except for negative feelings ($p = 0.16$). The domains where hypoglycemic events

Table 1 Internal consistency reliability at baseline

Domain	Cronbach's alpha
Usual activities	0.910
Pain/discomfort	0.656
Anxiety/depression	0.847
Satisfaction	0.688
Negative feelings	0.735
Positive feelings[a]	N/A
Feeling troubled	0.756

[a]Positive feelings only has 2 items, and therefore, alpha cannot be calculated

Table 2 Scale–scale correlations (Pearson) and significance level at baseline

Domain	UA	PD	AD	SAT	NEG	POS	FT
UA	1						
Usual activities							
PD	0.637	1					
Pain/discomfort	< 0.0001						
AD	0.645	0.638	1				
Anxiety/depression	< 0.0001	< 0.0001					
SAT	0.508	0.494	0.650	1			
Satisfaction	< 0.0001	< 0.0001	< 0.0001				
NEG	0.660	0.519	0.719	0.673	1		
Negative feelings	< 0.0001	< 0.0001	< 0.0001	< 0.0001			
POS	0.374	0.300	0.511	0.657	0.469	1	
Positive feelings	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001		
FT	0.574	0.760	0.653	0.404	0.459	0.296	1
Feeling troubled	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	

Table 3 Scale descriptive statistics at baseline

Domain	N	Mean	SD	Median	Minimum	Maximum	% min.	% max.	% missing
Usual activities	285	82.16	18.92	87.04	5.56	100	0	26	0
Pain/discomfort	285	82.72	18.39	87.50	20.83	100	0	30	0
Anxiety/depression	285	68.46	24.94	70.00	0.00	100	1	16	0
Satisfaction	285	66.47	23.05	66.67	0.00	100	1	14	0
Negative feelings	285	70.12	24.87	72.22	0.00	100	1	20	0
Positive feelings	285	65.82	22.90	66.67	0.00	100	1	18	0
Feeling troubled	285	80.94	23.34	88.89	5.56	100	0	40	0

Table 4 Scale descriptive statistics at retest ($n=91$)

Domain	N	Mean	Std Dev	Median	Minimum	Maximum	% min.	% max.	% missing
Usual activities	90	82.88	19.88	87.04	16.67	100	0	31	1
Pain/discomfort	90	83.24	19.23	89.58	29.17	100	0	34	1
Anxiety/depression	90	69.56	25.58	71.67	0.00	100	2	15	1
Satisfaction	88	67.58	22.16	66.67	0.00	100	1	8	3
Negative feelings	90	70.99	26.66	77.78	0.00	100	1	22	1
Positive feelings	88	64.77	23.43	62.50	0.00	100	3	13	3
Feeling troubled	91	84.55	21.80	94.44	16.67	100	0	46	0

Table 5 Clinical validity: hypoglycemic event occurrence in the last month

Hypoglycemic event in the last month	N	UA		PD		AD		SAT		NEG		POS		FT	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hypo event	65	73.36	21.31	69.94	16.54	58.26	26.07	58.55	25.59	66.32	24.96	60.90	24.29	62.74	26.78
No hypo noted	220	84.76	17.37	86.50	17.20	71.47	23.83	68.81	21.75	71.24	24.78	67.27	22.32	86.31	19.23
ANOVA p value		<0.0001		<0.0001		0.0002		0.0015		0.1621		0.0484		<0.0001	

UA usual activities, *PD* pain/discomfort, *AD* anxiety/depression, *SAT* satisfaction, *NEG* negative feelings, *POS* positive feelings, *FT* feeling troubled

were thought to manifest (usual activities, pain/discomfort, anxiety/depression, and feeling troubled) were all highly significant.

Known-group validity of the DTR-QOL domains was assessed using the patient's assessment of their glycemic control in the last month, concerns with weight gain in the last month, their current health status, and the degree of communication that they have with their clinician (Table 6). Most domains showed monotonic relationships with domain scores at each level of response, demonstrating that the domains are responding as expected. For glycemic control, all domains except pain/discomfort, positive feelings, and feeling troubled had monotonically decreasing scores for all responses. Evaluating the top 3 response options found that all scales were monotonically decreasing. Domains were not able to differentiate well with regard to a patient's concerns with weight gain over the last month. Only negative feelings and feeling troubled demonstrated monotonic, decreasing scores with greater concern. Small samples for some response options may be driving these results.

Similarly, the patient's perceived health status was not able to demonstrate a monotonic, decreasing pattern for domain scores except for anxiety/depression and negative and positive feelings. Evaluating the top three response options found that all scales were monotonically decreasing. The degree to which patients communicate with their clinician demonstrated a monotonically decreasing domain scores in the top three response options (too few to evaluate in the lower response options) in all domains except feeling trouble, which showed little differentiation between any of the response options. It should be noted that the items in feeling troubled are directly related to how "troubled" a patient feels with their current diabetic health state, something that is likely not affected by discussions with a clinician.

Overall, the DTR-QOL domains were able to differentiate across response levels of these known-group assessments. Given the small sample sizes in several response options, it was recommended to reassess known-group validity in future studies to further document its properties.

Table 6 Known-group validity

	N	UA		PD		AD		SAT		NEG		POS		FT	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Glycemic control in the last month															
1 Very good	30	86.23	18.04	85.14	18.66	78.00	24.43	81.48	20.75	81.48	19.43	81.67	21.82	86.85	21.07
2 Good	146	84.58	16.73	84.93	18.17	73.77	23.76	73.78	21.53	75.04	23.73	71.35	22.70	83.18	24.07
3 Neither good nor poor	80	78.52	20.94	78.54	18.75	61.58	23.15	55.35	17.67	65.90	22.81	54.17	17.54	77.50	21.92
4 Poor	24	78.01	21.72	81.60	17.55	53.06	22.46	47.92	14.73	48.84	23.57	52.43	18.47	72.22	24.95
5 Very poor	4	58.80	19.09	78.13	13.77	50.00	19.44	20.83	12.32	23.61	6.99	62.50	15.96	80.56	14.70
ANOVA *p* value		0.0068		0.1304		< 0.0001		< 0.0001		< 0.0001		< 0.0001		0.0786	
Concerned with weight gain in last month															
1 Not concerned at all	117	85.83	17.15	87.22	16.37	75.98	23.07	72.74	21.15	77.54	22.79	71.79	21.83	88.51	18.47
2 Not very concerned	75	80.49	18.83	80.00	18.37	69.38	24.71	66.37	22.61	72.15	21.58	65.67	22.30	78.81	24.37
3 Maybe a bit concerned	32	83.10	14.47	83.98	14.28	58.96	22.58	57.29	21.65	63.37	26.81	60.68	24.43	78.13	24.07
4 Somewhat concerned	50	77.30	22.55	78.33	20.72	58.33	23.83	59.00	24.30	57.56	25.69	56.83	22.57	72.11	24.11
5 Very concerned	9	69.14	25.07	68.06	28.49	58.52	26.09	59.26	29.66	50.00	24.85	62.04	20.46	59.88	33.65
ANOVA *p*-value		0.0126		0.0015		< 0.0001		0.0004		< 0.0001		0.0013		< 0.0001	
Health status															
1 Good	86	88.70	14.75	90.50	12.65	81.24	21.43	77.33	22.14	81.59	20.17	74.81	23.74	90.05	18.30
2 Somewhat good	108	82.53	19.49	82.60	18.04	68.40	24.52	67.64	21.45	71.35	23.92	67.67	20.94	80.86	22.47
3 Neither good nor poor	57	75.57	18.87	74.78	20.40	58.07	22.55	53.51	20.83	58.58	23.78	56.58	22.37	71.25	25.22
4 Somewhat poor	32	76.45	20.12	78.26	20.13	56.46	20.42	57.64	19.49	57.99	26.25	53.13	15.95	76.56	24.75
5 Poor	1	29.63	.	37.50	.	20.00	.	27.78	.	22.22	.	41.67	.	16.67	.
ANOVA *p*-value		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001		< 0.0001	
Degree of communication with clinician															
1 Good communication	157	84.34	18.28	84.34	16.79	72.31	24.20	71.73	22.94	72.82	24.97	70.91	23.24	83.37	21.33
2 Some communication	108	80.26	19.53	81.52	20.17	65.71	24.61	61.01	21.12	67.49	25.13	61.34	20.52	78.81	25.36
3 Cannot say one way or the other	13	76.35	16.84	80.77	17.72	60.26	24.47	54.70	20.27	62.39	19.13	48.72	14.77	79.06	20.56
4 Not much communication	5	72.59	26.10	72.50	24.04	45.33	19.94	56.67	36.09	70.00	26.53	53.33	37.55	63.33	39.01
5 No communication	1	59.26	.	50.00	.	50.00	.	33.33	.	55.56	.	50.00	.	61.11	.
ANOVA *p*-value		0.1296		0.1625		0.0199		0.0003		0.3251		0.0002		0.1815	

UA usual activities, *PD* pain/discomfort, *AD* anxiety/depression, *SAT* satisfaction, *NEG* negative feelings, *POS* positive feelings, *FT* feeling troubled

Concurrent validity was assessed by comparing DTR-QOL domain scores to the DTSQ domain scores and the SF-8 PCS and MCS (Table 7). Moderate correlations were found among similar domains, as expected. The DTSQ Treatment Satisfaction Score was most highly correlated with the satisfaction domain ($r=0.56$), positive feelings domain ($r=0.52$), and anxiety/depression ($r=0.52$). Negative events such as frequency of hyperglycemia and hypoglycemia demonstrated negative correlations across all domains. Perceived frequency of hyperglycemia demonstrated the strongest relationship with Satisfaction ($r=-0.52$). In contrast, perceived frequency of hypoglycemia had its strongest association with Pain/Discomfort ($r=-0.35$). The SF-8 summary scores demonstrated small-to-moderate correlations across DTR-QOL domains. The PCS and MCS had their strongest association with satisfaction ($r=0.32$) and anxiety/depression ($r=0.39$), respectively.

Test–retest reliability was found to be between 0.91 (Usual Activities) and 0.62 (positive feelings; Table 8). All scales except positive feelings met acceptable levels of reliability (0.70). Test–retest reliability was assessed on the entire population of patients, whose symptoms and treatment course were judged to be stable. It was recommended that test–retest reliability was re-evaluated in future studies, where a Patient Global Impression of Change (PGI-C) or some other measure of patient assessed stability is used to ensure patients truly are unchanged.

Table 8 Test–retest reliability

Domain	Shrout–Fleiss reliability (2,1)
Usual activities	0.912
Pain/discomfort	0.811
Anxiety/depression	0.876
Satisfaction	0.741
Negative feelings	0.897
Positive feelings	0.620
Feeling troubled	0.840

Discussion

The DTR-QOL was reconstructed into seven, more unidimensional domains based on item concepts followed by psychometric evaluation. Although several items are broad and are highly correlated with other items, the overall fit of the new structure was acceptable.

Although there remained two mismatched items with respect to item-scale correlations, the item content dictated that they remained in their current domain to ensure a higher level of interpretation of the scores. Internal consistency reliability was also less than desired for pain/discomfort and satisfaction domains. Satisfaction had

Table 7 Concurrent validation at baseline

Concurrent Measure	UA	PD	DTR-QOL domains			POS	FT
			AD	SAT	NEG		
DTSQ Treatment Satisfaction Score (0–36)							
correlation ($r=$)	0.460	0.375	0.516	0.559	0.401	0.523	0.275
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
number of patients ($n=$)	283	283	283	283	283	283	283
DTSQ perceived freq of hyperglycemia (0–6)							
correlation ($r=$)	−0.346	−0.343	−0.420	−0.517	−0.467	−0.393	−0.278
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
number of patients ($n=$)	285	285	285	285	285	285	285
DTSQ perceived freq of hypoglycemia (0–6)							
correlation ($r=$)	−0.260	−0.347	−0.267	−0.195	−0.135	−0.149	−0.341
p-value	<0.0001	<0.0001	<0.0001	0.0009	0.0228	0.0118	<0.0001
number of patients ($n=$)	285	285	285	285	285	285	285
SF-8 physical component summary (PCS)[a]							
correlation ($r=$)	0.261	0.298	0.304	0.319	0.315	0.253	0.272
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
number of patients ($n=$)	284	284	284	284	284	284	284
SF-8 mental component summary (MCS)*							
correlation ($r=$)	0.358	0.350	0.391	0.278	0.362	0.157	0.239
p-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0081	<0.0001
number of patients ($n=$)	284	284	284	284	284	284	284

[a]PCS/MCS scored with a mean of 50 and standard deviation of 10

a documented alpha of 0.69, which was borderline on acceptable (using a 0.70 threshold). This lower reliability will require larger changes to demonstrate significance and may be found to be higher in another more diverse population.

There were some concerns with ceiling effects; however, this needed to be considered in terms of the population under investigation. Using a more diverse population with more severe patients will likely demonstrate fewer ceiling effects. The results of the clinical validity testing on "hypoglycemic event in the last month" demonstrated the ability of the DTR-QOL domains to differentiate when a clinical event occurred.

Concurrent validity was established, with domains measuring similar concepts more highly correlated than domains of different concepts. Finally, test–retest were found to be acceptable for all domains except positive feelings. A limitation of this analysis was the lack of a marker for patient assessed "unchanged". Re-evaluating test–retest on a documented stable population may show better properties. In addition, evaluation of the DTR-QOL on patients through time would be beneficial to evaluate the instrument's ability to detect change.

Conclusions

The restructuring of the DTR-QOL resulted in an acceptable measure with domains that are more easily interpreted and will allow physicians to better measure more specific impacts on their patients. It is recommended to replicate the validation analyses on a population with a larger range of severity and complications to ensure a diverse population. This would allow the evaluation of statistical properties within a population, where differences are expected and provide further validation supporting the use of the DTR-QOL. A longitudinal analysis is also recommended to evaluate the DTR-QOL's usefulness in detecting changes and in monitoring treatment.

Acknowledgements The authors would like to thank Ayumi Tomari for her assistance in developing the new structure of the DTR-QOL based on conceptual meanings.

Author contributions All authors were involved in the restructuring of the DTR-QOL, interpretation of the analysis, and manuscript development, and BC also conducted the analysis.

Funding Novo Nordisk Pharma Ltd provided financial support for this study.

Compliance with ethical standards

Conflict of interest The authors declare no other conflicts, funding, or other interests.

Human rights and informed consent This study does not involve human subjects as it is a secondary use of a data set provided by the original validation study.

References

1. Gæde P, et al. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. N Engl J Med. 2003;348(5):383–93.
2. The Ministry of Health, L.a.W. Patient survey 2014. 2014. http://www.mhlw.go.jp/toukei/saikin/hw/kanja/14/index.html. Accessed 31 Oct 2017.
3. The Ministry of Health, L.a.W. Outline of the results of the National Health and Nutrition Survey. 2016. http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/kekkagaiyou_7.pdf. Accessed 31 Oct 2017.
4. de Climens AR, et al. Review of patient-reported outcome instruments measuring health-related quality of life and satisfaction in patients with type 2 diabetes treated with oral therapy. Curr Med Res Opin. 2015;31(4):643–65.
5. Ishii H, et al. Improvement of glycemic control and quality-of-life by insulin lispro therapy: assessing benefits by ITR-QOL questionnaires. Diabetes Res Clin Pract. 2008;81(2):169–78.
6. Polonsky WH, Henry RR. Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors. Patient Prefer Adherence. 2016;10:1299–307.
7. Cramer JA. A systematic review of adherence with medications for diabetes. Diabetes Care. 2004;27(5):1218–24.
8. Biderman A, et al. Treatment satisfaction of diabetic patients: what are the contributing factors? Fam Pract. 2009;26(2):102–8.
9. Ishii H. Development and psychometric validation of the diabetes therapy-related QOL (DTR-QOL) questionnaire. J Med Econ. 2012;15(3):556–63.
10. Rothman M, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO Task force report. Value Health. 2009;12(8):1075–83.
11. FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79.
12. Patrick DL, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health. 2007;10(Suppl 2):S125–37.
13. Ishii H, et al. Improvement of quality of life through glycemic control by liraglutide, a GLP-1 analog, in insulin-naive patients with type 2 diabetes mellitus: the PAGE1 study. Diabetol Metab Syndr. 2017;9:3.
14. Mashitani T, et al. Diabetes treatment-related quality of life is associated with levels of self-care activities in insulin injection among Japanese patients with type 2 diabetes: diabetes distress and care registry at Tenri (DDCRT 8). Acta Diabetol. 2015;52(4):639–47.
15. Okada M, et al. Effect of switching basal insulin regimen to degludec on quality of life in Japanese patients with type 1 and type 2 diabetes mellitus. J Pharm Health Care Sci. 2015;1:26.
16. Ishii H, et al. The Japanese version of the diabetes treatment satisfaction questionnaire (DTSQ): translation and clinical evaluation. J Clin Exp Med. 2000;192(7):809–14.
17. Bradley C. The diabetes treatment satisfaction questionnaire: DTSQ. In: Handbook of psychology and diabetes: a guide to

psychological measurement in diabetes research and practice; 1994, pp. 111–32.
18. Fukuhara S, Suzukamo Y. Manual of the SF-8 Japanese version. Kyoto: Institute for Health Outcomes & Process Evaluation Research; 2004. p. 71–116.
19. Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the food and drug administration and the reflection paper of the European medicines agency. Eur J Cancer. 2009;45(3):347–53.
20. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52(4):281–302.