BMJ Open Diabetes Research & Care

Psychometric testing of the Norwegian Diabetes Health Profile (DHP-18) in patients with type 1 diabetes

Lars-Petter Jelsness-Jørgensen,[1,2] Øystein Jensen,[1,3] Charlotte Gibbs,[4] Ragnar Bekkhus Moe,[2] Dag Hofsø,[5] Tomm Bernklev[3,6]

ABSTRACT

Objective The Diabetes Health Profile-18 (DHP-18) was developed to measure disease-specific health-related quality of life. It has been translated into Norwegian but remains invalidated. The purpose of this paper was to examine the psychometric properties of the Norwegian DHP-18.

Research design and methods Participants with type 1 diabetes were recruited from three outpatient clinics in Norway. Clinical and sociodemographic data were collected, and participants completed the DHP-18 and the Short-Form 36 (SF-36). Descriptive analysis, frequencies, t-tests and the chi-squared tests were used. Principal axis factoring (PAF) and confirmatory factor analysis (CFA) were used. Convergent validity was tested using Spearman's correlation between the DHP-18 and SF-36. Reliability was tested using Cronbach's alpha and intraclass correlation coefficient.

Results In total, 288 patients were included. No floor and ceiling effects were found. A forced PAF analysis revealed that three questions had an eigenvalue below 0.40. In the unforced PAF analysis, one question loaded below 0.40, while three questions loaded into a fourth factor. The correlation between the DHP-18 and SF-36 dimensions was low to moderate. Problematic internal consistency was observed for the disinhibited eating dimension in the forced PAF and in the suggested fourth dimension in the unforced PAF. CFA revealed poor fit. The test–retest reliability displayed good to excellent values, but responsiveness was limited.

Conclusions Problematic issues were identified regarding factor structure, item loadings, internal consistency and responsiveness. Further evaluation of responsiveness is particularly recommended, and using a revised 14-item DHP version is suggested.

Significance of this study

What is already known about this subject?
► Health-related quality of life is impaired in type 1 diabetes compared with healthy controls.
► The Diabetes Health Profile (DHP) has been developed to assess the psychological and behavioral burden of living with diabetes.

What are the new findings?
► This is the first psychometrical testing of the DHP-18 in Norwegian sample of patients with type 1 diabetes.
► Problematic issues were identified regarding factor structure, item loadings, internal consistency and responsiveness.

How might these results change the focus of research or clinical practice?
► A revised 14-item DHP version is suggested.
► Further clarification on psychometrical properties is needed before implementation in clinical practice and studies.

INTRODUCTION

Being diagnosed with type 1 diabetes (T1D) may affect patients negatively across their lifespan, and several studies have shown that T1D is associated with impaired health-related quality of life (HRQoL).[1]

The term HRQoL was introduced to distinguish between quality of life in a more general sense and the requirements of clinical medicine and clinical trials.[2] However, HRQoL continues to have a quite unclear definition. In general it is agreed that relevant aspects may vary from study to study, but may include aspects such as physical, psychological, social, and emotional functioning, as well as general health, symptoms and existential issues.[2] Even though some instruments may focus on a single concept, there is a consensus that a number of the above dimensions should be included in HRQoL questionnaires. In general, we distinguish between two main types of instruments used to measure HRQoL in clinical populations. While generic instruments may be used in the general population and across disease groups, disease-specific instruments have been developed to detect more subtle disease and treatment-related effects.[2] Disease-specific and generic instruments have a complimentary relationship. For example, generic measures can monitor changes in the physical functioning of patients over time in relation

to population norms, regardless of the cause of change. Disease-specific measures can help determine which conditions accounted most for a patient's limitations in physical functioning and, therefore, make such measures more useful in outcomes research, studies of healthcare costs, and clinical practice.[3]

The Diabetes Health Profile-18 (DHP-18) was developed with significant patient and clinical input to adequately assess the psychological and behavioral burden of living with diabetes.[4] Because of the difficulties and complexities of defining and measuring the patient's quality of life, the focus of the original DHP (DHP-1) was the psychological and behavioral dysfunctioning of the person as a consequence of the impact of living daily with diabetes. Therefore, it was considered important that the measure contained content that reflected the outcome from the everyday dynamic interchange between the person with diabetes and the environment and as a result dysfunctional outcomes could be addressed or alleviated through appropriate educational or therapeutic intervention.[5] The 'Transactional Theory of Stress and Coping' was used as the theoretical model underpinning the development of the DHP-1.[6] The rationale for the development of the DHP-18 was the same as the DHP-1 with the difference to develop an instrument using questions from the DHP-1 question set that would be suitable for use across both type 1 and type 2. The questionnaire has consequently undergone psychometrical testing in both T1D and type 2 diabetes (T2D).[7]

In order to be used in clinical settings and studies, psychometrical testing is an essential methodological process, that is, testing of reliability, validity and sensitivity to change. While validity concerns whether an instrument measures what it is intended to measure, reliability concerns if the instrument at hand yields reproducible results over time under consistent conditions. Sensitivity to change is a measure of the ability of an instrument to detect clinically relevant differences, for example, change in health status.[2] Previous research has demonstrated that the DHP-18 has displayed high levels of validity, reliability and patient acceptability,[4] but the overall sensitivity to change has been found to be limited.[8] However, the latter may be related to the way in which change was measured, namely by using a generic self-reporting question. Thus, potentially limiting the ability to detect relevant changes.[8] However, even though an instrument has displayed acceptable psychometric properties in one language, one cannot automatically conclude that the same will apply in other languages and cultures. Cross-cultural adaptation is consequently important.[9] No data have been published regarding the validity, reliability and sensitivity to change of the Norwegian DHP-18.[2] Hence, the aim of this study was to test the psychometric properties of the Norwegian version of the DHP-18.

RESEARCH DESIGN AND METHODS
Patients
In this cross-sectional, descriptive study, patients diagnosed with T1D, 18 years or older, were consecutively included during routine follow-up at three diabetes outpatient clinics in the southeastern part of Norway from May 2015 to November 2016. Sociodemographic, clinical and laboratory tests were collected at baseline. Sociodemographic data included age, gender, civil status, level of education (dichotomized into lower and higher levels of education (ie, education beyond the secondary level)), work status (either working or not working) and smoking status (dichotomized into current and former smokers in one group, non-smokers in the other). Clinical data included body mass index, disease duration, medication use, comorbidity, diabetes-related complications and the Wagner classification of foot ulcers. Laboratory data included hemoglobin, C-reactive protein, ferritin, vitamin D (25-hydroxy), leukocytes and iron.

To examine the validity, reliability and sensitivity to change of the DHP-18, all patients were asked to fill out the questionnaires at inclusion (baseline) and after 4–6 weeks (retest). At retest, patients also completed a question concerning their subjective health state; 'Compared to the last time you completed the questionnaire, how do you evaluate your condition today? (i) unchanged, (ii) improved, or (iii) deteriorated.' The questionnaires were sent by mail, and patients were asked to complete and return them in a prestamped envelope. At each center, a senior endocrinologist was responsible for study performance (ie, served as the local principal investigator).

Data collection
Questionnaires
Diabetes Health Profile-18
The DHP-18[4] consists of 18 questions assessing psychological and behavioral functioning, divided into three dimensions: psychological distress (six questions), barriers to activity (seven questions) and disinhibited eating (five questions). The scoring method, which is applied to the DHP, is based on the widely used Likert method of summated scales in which each question is scored using a graded scale and summated to provide a total score for the specific domain. Each of the 18 questions is consequently scored on a 0–3 scale, and transformed into a 0–100 scale, where a higher score indicates lower levels of HRQoL.[8] For the DHP-18, a number of different 'forced choice' adjective scales are used to measure either frequency or intensity which depend on the nature of the question asked. The DHP-18 has been translated and linguistically validated into Norwegian according to the principles of good practice for the translation and cultural adaptation process for patient-reported outcome measures (PROM).[10]

Short-Form 36
The Short-Form 36 (SF-36) is a generic HRQoL questionnaire designed to assess functional status, well-being,

and general perception of health.[11] The questionnaire consists of 36 questions, which are transformed into eight dimensions: physical functioning (10 questions), bodily pain (2 questions), vitality or energy level (4 questions), social functioning (2 questions), mental health (5 questions), general health (5 questions), role limitation due to physical problems (4 questions), and role limitations due to personal or emotional problems (3 questions). An additional question reports on health transition over the past year. For each question, the raw score was coded and transformed into a scale from 0 to 100, with 0 indicating the lowest level of function and 100 the highest level of function. The questionnaire has been translated into Norwegian[12] and has been validated among people with diabetes.[13]

Procedure

A standardized inclusion procedure was followed at each center. This included baseline collection of sociodemographic, clinical, laboratory and patient-reported outcome data. Moreover, this procedure enabled patients to fill out the questionnaires undisturbed at the hospital outpatient clinic. While clinical data were collected during clinical consultation with an endocrinologist, sociodemographic data were self-reported by patients. Moreover, laboratory data were based on blood samples drawn by a phlebotomist in connection with the clinical consultation. All questionnaires were collected and checked by a study nurse before the patient left the consultation to optimize data completeness and quality. Face validity was investigated by distributing the questionnaire to 10 patients prior to the main study. This was done to receive patient input on question content, scoring, and structure. Missing values for the DHP-18 and SF-36 were treated as recommended by Meadows[5] and Ware[14]: if ≥50% of questions in a dimension had been completed, missing values were substituted with the mean of the completed questions for that dimension.

Statistical analysis

To assess the characteristics of the sample, we used descriptive analysis, frequencies, t-tests and the χ^2 test. Floor and ceiling effects were investigated by calculating the percentage of patients scoring either the lowest or highest possible dimensional scores. A 15% cut-off value for both floor and ceiling effects was used according to recommendations in the literature.[15] The construct of the Norwegian DHP-18 was tested with principal axis factoring (PAF) analyses with varimax orthogonal rotation. In accordance with the original validation studies, both forced and unforced PAF were used.[4 7] In addition, a confirmatory factor analysis (CFA) was performed to investigate the model fit.

Construct validity was further tested using three approaches: (1) convergent validity, (2) discriminant validity, and (3) known-group validity.

1. Convergent validity was calculated using binary correlation analysis (Spearman's r—due to evidence of non-normal value distributions) of the DHP-18 and SF-36. Before starting the analysis, we set up the following a priori hypothesis:
   – Based on semantic construct, we hypothesized that the DHP-18 dimension psychological distress would correlate with the SF-36 dimensions mental health, vitality, social functioning, general health and role emotional.
   – The DHP-18 dimensions barriers to activity and disinhibited eating were hypothesized to display low correlations with all SF-36 dimensions.
2. Discriminant validity was calculated by comparing the correlation between the three DHP-18 dimensions.
3. Student's t-test was used to calculate known-group validity by comparing DHP-18 scores in patients with or without diabetes-related complications and comorbidity. DHP-18 scores were also investigated in three groups: (A) no complications, (B) one to two complications, and (C) ≥3 complications. The known-group validation was based on the principle that certain specified groups of patients might be anticipated to score differently from others. Thus, the instrument should be sensitive to these differences.

Internal consistency reliability was calculated using Cronbach's alpha coefficient, where values above 0.7 are regarded as acceptable.[2] Test–retest reliability was calculated using the intraclass correlation coefficient (ICC) in patients reporting that their disease state was unchanged from baseline to retest (4 weeks' interval). Values from 0.70 to 0.90 represent 'moderate or good reliability' and above 0.90 'high or excellent'.[2] Responsiveness was tested in those patients who reported either deterioration or improvement in DHP-18 scores from baseline to retest, by using paired t-tests. Cohen's d effect size was used to investigate responsiveness and calculated by comparing the mean difference between groups, divided by the pooled SD. Operational definitions of 0.2, 0.5, and 0.8 were categorized as small, medium, and large, respectively.[16]

All tests were two sided, with a 5% significance level. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS V.24, IBM) and IBM AMOS V.25.

RESULTS

In total, 332 eligible patients were invited to participate in the study. Of these, 288 (87%) patients gave written informed consent. Baseline characteristics of the included patients are presented in table 1. Except for comorbidity (p=0.002), no statistically significant differences were observed between genders. A total of 199 patients completed the retest, corresponding to 69% of the original sample. Two patients did not complete the health condition question during the retest and were consequently excluded, leaving 197 patients with complete datasets. Compared with responders at the retest, non-responders were significantly younger (p<0.001), and were

Table 1 Sociodemographic and clinical data of included patients (n=288)

Age, mean (SD), years	44.8 (13.3)
Age range (min-max)	18–80
Civil status (n=) (missing, n=1)	
Single	38
Married/cohabitant	231
Divorced/separated	15
Widow/widower	3
Educational level (n=)	
Lower education	157
Higher education	131
Employment (n=) (missing, n=1)	
Working	202
Not working	85
Smoking (yes) (missing, n=2)	68
Diabetes duration, mean (SD)	23.0 (13.3)
Diabetes duration, range (min-max)	0–63
Insulin pump users (n=)	112
BMI, mean (SD)	26.7 (4.6)
Diabetes complications (n=)	
Neuropathy	38
Retinopathy	85
Nephropathy	20
Comorbidity mr_and sr_ yes (n=)	177
Plasma glucose*, mean (SD) (missing, n=2)	9.86 (4.7)
HbA1C, mean (SD) (missing, n=2)	8.16% (1.2) 66 mmol/mol

Figures presented as n= if not otherwise noted, lower education, =<university/college; higher education, college or university; plasma glucose=not fasting,
BMI, body mass index; mr, medical records; sr, self-reported.

more often male (p<0.001). There were no patients who exceeded the limit of 50% missing values in DHP-18, but six patients were excluded due to missing data in SF-36. No significant floor or ceiling effects were observed.

Validity
Face validity revealed no problematic issues regarding either content or scoring. At baseline and retest, 0.3% of patients made specific comments on single items in the DHP-18 (ie, items 2, 3, 4, 8, 9, 12, 14, 15, 16 and 17). Specific comments were related to questions of relevance and meaning. For example, in question 4 several commented on the meaning of 'avoid going out', which in Norwegian can imply anything from going out of the house to going to a party. Furthermore, in question 14 'Do you get edgy when out and there is nowhere to eat?', some commented that there is a huge difference between getting edgy because you cannot get hold food fast enough and getting edgy because you cannot find an available table at a restaurant.

Factor loadings of the forced PAF analysis accounted for 42% of the variance and are presented in table 2. Three questions (4, 13 and 14), all belonging to the original barriers to activity dimension, loaded below the cut-off of 0.40. The unforced PAF analysis explained 46% of the variance and is presented in table 3. One question (question 4) loaded below the cut-off of 0.40, while three questions from the original barriers to activity dimension (11, 13 and 14) loaded into a fourth factor. CFA displayed an χ^2 (132)=329.91, p<0.001, the values of confirmatory fit index=0.885 and root mean square error of approximation=0.071 indicated a poor fit to the data. The items did not load strongly on their respective factors and did not fit to the standardized $\lambda \geq 0.60$; all items were p<0.001, and indicated that the latent variables were not well defined by their items. The estimates for psychological distress ranged from 0.12 to 40, the estimates for barriers to activity ranged from 0.28 to 0.60, and the estimates for disinhibited eating ranged from 0.22 to 0.69. Correlations among the three latent variables ranged from (0.02) to (0.12).

Convergent validity between the DHP-18 dimensions and the SF-36 dimensions revealed, in accordance with the a priori hypotheses, an inverse relationship with all SF-36 dimensions (table 4). *Discriminant validity* showed an overall low correlation between the DHP-18 dimensions (0.22–0.37). *Known-group* validation revealed that the mean psychological distress was 24.2 vs 21.9 (p=0.02) for those with or without complications/comorbidity, respectively. No statistically significant differences were observed for the barriers to activity and disinhibited eating dimensions. When investigating DHP-18 scores according to whether patients had no, 1–2 or ≥3 complications, no significant differences were observed.

Reliability and responsiveness
Internal consistency reliability revealed an overall Cronbach's alpha of 0.79, while dimensional alphas were 0.85 for psychological distress, 0.70 for barriers to activity, and 0.42 for disinhibited eating. When investigating item-total statistics, the results indicated that excluding question 9 from the disinhibited eating dimension would increase the dimensional and overall Cronbach's alpha to 0.78 and 0.83, respectively. When excluding questions 4, 13 and 14 (all belonging to the barriers to activity dimension), which loaded <0.40, the overall alpha for the remaining 15 questions was 0.77, and the dimensional alpha remained 0.70. When investigating alpha values in the fourth factor suggested in the unforced PAF, Cronbach's alpha was 0.52 (questions 11, 13 and 14). The overall alpha after excluding questions 4, 9, 13 and 14 in the remaining 14 questions was 0.81.

A total of 156/197 (79.2%) reported that their condition was unchanged from baseline to retest, and ICC values were 0.82, 0.76 and 0.74 for psychological distress, barriers to activity and disinhibited eating, respectively.

There were a low number of patients reporting either improvement (20/197, 10.2%) or deterioration (21/197,

Table 2 Forced principal axis factoring (PAF) with varimax orthogonal rotation with Kaiser normalization

Dimensions and questions	Factor			Mean (SD)	Communalities
	1	2	3		
Psychological distress					
6. Do you lose your temper if people keep on at you about sugar testing or diet?	**0.537**	0.252	0.005	0.7 (0.8)	0.31
8. Are there more arguments or upsets at home than there would be if you did not have diabetes?	**0.568**	0.000	0.163	0.5 (0.7)	0.37
15. Because of your diabetes, do you get depressed?	**0.478**	0.308	0.316	0.8 (0.7)	0.43
16. Does your diabetes cause you to lose your temper or shout?	**0.835**	0.137	0.105	0.8 (0.7)	0.73
17. Do you get touchy or moody about diabetes?	**0.812**	0.177	0.187	0.8 (0.7)	0.73
18. Do you find yourself losing your temper over small things?	**0.720**	0.210	0.132	0.6 (0.7)	0.58
Barriers to activity					
1. Does food control your life?	0.016	0.159	**0.588**	1.3 (0.9)	0.40
2. Does having diabetes mean it is difficult staying out late?	0.139	−0.027	**0.573**	0.4 (0.7)	0.35
3. Does having diabetes mean your days are tied to meal times?	−0.007	−0.054	**0.756**	1.1 (1.0)	0.59
4. Do you avoid going out if your sugar is on the low side?	0.119	−0.058	0.378	0.8 (0.9)	0.17
11. Because of your diabetes, do you worry about getting colds or influenza?	0.139	−0.015	**0.505**	0.4 (0.7)	0.26
13. Do you find it frightening or worrying going into busy or crowded shops?	0.211	0.269	0.219	0.3 (0.6)	0.17
14. Do you get edgy when out and there is nowhere to eat?	0.260	0.179	0.354	0.6 (0.7)	0.24
Disinhibited eating					
5. Do you have problems keeping to your diet because you eat to cheer yourself up?	0.153	**0.752**	0.041	0.6 (0.8)	0.61
7. Do you have problems keeping to your diet because you find it hard to say no to food you like?	0.114	**0.710**	−0.018	0.9 (0.8)	0.54
9. When you start eating how easy do you find it to stop?	−0.112	−0.608	0.069	2.0 (0.7)	0.40
10. How likely are you to eat something extra when you feel bored or fed up?	0.164	**0.705**	−0.060	1.5 (0.8)	0.53
12. Do you wish there were not so many nice things to eat?	0.112	**0.553**	0.118	0.9 (1.0)	0.36

Values higher than 0.4 in boldface.

10.7%). No statistically significant differences were observed in any of the dimensional scores between baseline and retest. In addition, the effect sizes measured with Cohen's d were either undetectable or small. Responsiveness is presented in table 5.

DISCUSSION

In the present study, we investigated the psychometrical properties of the Norwegian version of the DHP-18. Even though satisfactory psychometric properties were observed in a substantial number of aspects, problematic issues were identified regarding factor structure, item loadings, internal consistency, and sensitivity to change.

Even though we did not observe any problematic issues when testing for face validity, several critical comments were provided at both baseline and retest. The reason for this discrepancy is unclear but may potentially be related to the small sample of patients (n=10) participating in the face validity test. A plausible explanation for some of the critical comments made by patients may be that some questions are less relevant almost 20 years after the original development of the questionnaire. For instance,

improved technology such as continuous glucose monitoring (CGM) increases hypoglycemic confidence and decreases diabetes distress, and CGM contributes to significant improvement in diabetes-specific quality of life.[17] Diabetes distress refers to the worries, concerns, and fears that are relatively common among individuals who struggle with a progressive and demanding chronic disease, and high levels of diabetes distress have been linked to problematic diabetes management and poor glycemic control.[17 18]

The forced three-factor PAF analysis revealed that the factors consisted of the same single items as in the original study, but three questions (4, 13 and 14) loaded under the recommended value of 0.40.[19] Concerning question 4, our results align with the findings of Meadows et al.[7] Moreover, the fact that more of the questions loaded below 0.40 than in previous studies may be related to the different patient populations used.[7 20] While we merely investigated patients with T1D, Meadows et al[7] investigated patients with T2D, and Tan et al[20] studied a combination of patients with T1D and T2D. To explore if another factor structure might be observed in the

Table 3 Unforced principal axis factoring (PAF) with varimax orthogonal rotation with Kaiser normalization

Dimensions and items		Factor 1	2	3	4	Communalities
Psychological distress						
6.	Do you lose your temper if people keep on at you about sugar testing or diet?	**0.523**	0.250	−0.033	0.115	00.31
8.	Are there more arguments or upsets at home than there would be if you did not have diabetes?	**0.617**	0.007	0.187	−0.018	0.42
15.	Because of your diabetes, do you get depressed?	**0.427**	0.272	0.171	0.416	0.48
16.	Does your diabetes cause you to lose your temper or shout?	**0.819**	0.133	0.042	0.181	0.73
17.	Do you get touchy or moody about diabetes?	**0.779**	0.160	0.085	0.283	0.72
18.	Do you find yourself losing your temper over small things?	**0.728**	0.212	0.108	0.109	0.60
Barriers to activity						
1.	Does food control your life?	0.060	0.177	**0.699**	0.006	0.53
2.	Does having diabetes mean it is difficult staying out late?	0.157	−0.043	**0.536**	0.180	0.34
3.	Does having diabetes mean your days are tied to meal times?	0.024	−0.070	**0.750**	0.172	0.60
4.	Do you avoid going out if your sugar is on the low side?	0.089	−0.093	0.269	0.300	0.19
11.	Because of your diabetes, do you worry about getting colds orinfluenza?	0.088	−0.063	0.365	**0.432**	0.31
13.	Do you find it frightening or worrying going into busy or crowded shops?	0.110	0.227	0.015	**0.551**	0.37
14.	Do you get edgy when out and there is nowhere to eat?	0.201	0.139	0.210	**0.423**	0.28
Disinhibited eating						
5.	Do you have problems keeping to your diet because you eat to cheer yourself up?	0.135	**0.741**	0.011	0.141	0.61
7.	Do you have problems keeping to your diet because you find it hard to say no to food you like?	0.116	**0.714**	−0.002	0.028	0.54
9.	When you start eating, how easy do you find it to stop?	−0.115	−0.611	0.056	−0.008	0.40
10.	How likely are you to eat something extra when you feel bored or fed up?	0.156	**0.704**	−0.059	−0.056	0.53
12.	Do you wish there were not so many nice things to eat?	0.108	**0.547**	0.107	0.099	0.36

Values higher than 0.4 in boldface.

Norwegian validation, we also performed an unforced PAF analysis suggesting a four-factor solution. Question 4, however, remained below the threshold of 0.40. Moreover, results of the CFA indicated a poor fit to the data.

The strongest correlation was, as expected, between the psychological distress dimension in the DHP-18 and the mental health dimension in the SF-36. Similar findings have been reported in other studies.[4 7 21] As expected,

Table 4 Correlation (Spearman's r) between the Diabetes Health Profile-18 dimensions and the Short Form-36 dimensions

SF-36 dimensions	Diabetes Health Profile-18 Psychological distress	Barriers to activity	Disinhibited eating
Physical functioning	−0.244	−0.280	−0.207
Role physical	−0.304	−0.291	−0.174
Bodily pain	−0.243	−0.205	−0.167
General health	−0.440	−0.265	−0.259
Vitality	−0.423	−0.196	−0.333
Social functioning	−0.475	−0.297	−0.260
Role emotional	−0.422	−0.256	−0.216
Mental health	−0.527	−0.307	−0.257

All values are significant with p<0.005. Scores in the range 0.40–0.74 are in boldface.
SF-36, Short-Form 36.

Table 5 Diabetes Health Profile-18 subdimensions: psychological distress, barriers to activity, disinhibited eating and change in health condition from baseline to retest

Condition	Psychological distress				Barriers to activity				Disinhibited eating			
	Baseline	Retest	P values	Cohen's d	Baseline	Retest	P values	Cohen's d	Baseline	Retest	P values	Cohen's d
Improved, mean (SD)	25.6 (17.0)	27.2 (18.7)	0.545	–0.1	24.0 (12.8)	24.7 (13.8)	0.710	–0.05	38.0 (13.3)	41.3 (14.1)	0.205	–0.24
Deteriorated, mean (SD)	30.4 (18.3)	32.2 (15.0)	0.728	–0.1	26.4 (17.3)	29.3 (12.6)	0.433	–0.19	47.0 (14.6)	44.4 (14.4)	0.338	0.18

based on few common features, low to moderate correlations were observed between the DHP-18 dimension barriers to activity/disinhibited eating and all SF-36 dimensions. Of course, these findings also support discriminant ability and indicate a generally low level of overlap between the constructs of the SF-36 and the DHP-18. Furthermore, and in accordance with Meadows *et al*,[7] estimation of discriminant validity indicated low correlation between the DHP-18 dimensions.

Known-group validity was investigated by comparing DHP-18 scores in patients with or without comorbidity and complications. Even though our findings indicate that the DHP-18 can distinguish between these groups, we also observed some inconsistencies across the three dimensions. Similar findings are reported by Mulhern and Meadows.[8] A potential explanation may be that diabetes-related complications and comorbidity do not affect the disinhibited eating dimension to the same extent as the other two dimensions do. On the other hand, choosing a different parameter to investigate discriminant ability could potentially also have yielded a different result, for example, comparing newly diagnosed patients to those with an established diagnosis. Even though previous studies have reported that the DHP-18 can discriminate between different levels of illness, the fact that these results are based on patients with T2D limits direct comparison with the current study.[7 8] Of note, when dividing complications into three groups, no significant differences were observed.

Except for the disinhibited eating dimension, good internal consistency was found. The low internal consistency of the disinhibited eating dimension is in contrast to other studies[4 7 20] and may be related to the different patient populations investigated.[7 20] Based on a more detailed analysis of item-total statistics, we observed that the removal of question 9 increased the internal consistency to an acceptable level.[2] In addition, when investigating the dimensional alphas in the suggested four-factor solution, the internal consistency was low.[2]

The test–retest reliability showed moderate to excellent values in accordance with recommendations in the literature.[2 7] The sample size needed for test–retest analysis has been the subject of some debate.[2 22] Some have advocated that a sample size of 50 could be sufficient as a starting point,[23] while others have highlighted the need for larger sample sizes and more robust test–retest data.[24]

Hence, a strength of our study is the large sample size included in the retest analyses.

A central aspect of a PROM is the ability to respond to relevant changes in a particular condition, also known as sensitivity to change. Optimally, a PROM should be able to discriminate between groups of patients who report differences in health status. In this study, a marginal number reported either improvement or deterioration, increasing the risk of a type II statistical error. With this limitation in mind, we were not able to observe any statistically significant differences in any of the groups. Other studies have also reported low responsiveness.[8 25] The ability of a disease-specific questionnaire to capture relevant changes in a condition is critical, and future studies must, therefore, keep this in mind in order to clarify whether the DHP-18 indeed is responsive to change.

In addition to the factors discussed previously, it is of course hard to say whether or not different sample characteristics in the current and former studies may explain the findings in this study. While patients in the study by Meadows *et al*[4] were somewhat younger than our population, patients in the study by Mulhern and Meadows[8] were of higher age. However, the latter study investigated the validity and reliability of the DHP-18 in a cohort of patients with T2D, which consequently might explain the large age difference between the studies. Further, the level of diabetes-related complications did not differ between Mulhern and Meadows[8] and our study.

Based on the methodological observations made in this study, including that of factor structure, item loadings and internal consistency, we argue for excluding the following items: item 4 (loading below the recommended limit of 0.40 in both the forced and unforced PAF), and items 9, 13 and 14 (either loading under 0.40 or resulting in weak internal consistency). Therefore, using a 14-item version of the Norwegian DHP could be suggested (online supplementary appendix 1). However, such a choice is not without limitations, particularly since it hampers direct comparison to international studies that have used the original 18-item version of the DHP.

This study has some limitations in addition to those factors previously discussed. The psychometrical testing was not performed in T2D, consequently limiting the applicability of the results merely to T1D. We evaluated change in health status using the patient's own subjective experience. Using a more objective marker of disease

could have strengthen the analyses. Known-group validation was investigated by comparing patients with complications and comorbidities to those without. Of course, such a simplistic way of defining known groups may be viewed as a limitation. In addition, we do not have any information regarding those patients who were not included in the study, and even though a consecutive recruitment procedure was undertaken, we cannot exclude the risk of a potential recruitment bias. In our study, we used SF-36 to investigate criterion validity. Based on content in DHP-18 and SF-36 we expected low correlations in two of the DHP-18 dimensions compared with SF-36. In retrospect, another instrument than the SF-36 might have been more proper to measure criterion validation. We are also aware that other questionnaires could be used to measure the aspects that DHP-18 focuses on. However, the primary rationale for choosing the DHP-18 was to investigate the validity, reliability and sensitivity of this instrument in a Norwegian population of patients with T1D, since this had not been done previously. Further, if ≥50% of questions in a dimension had been completed, missing values were substituted with the mean of the completed questions for that dimension. We realize that use of the 'half rule' should be used with caution when questions have been ordered hierarchically. However, the DHP does not have a hierarchical structure to the ordering of its questions and therefore, using the 'half rule' is the specified method for substituting missing values for all versions of the DHP.

CONCLUSIONS

Problematic issues were identified regarding factor structure, item loadings, internal consistency and responsiveness of the Norwegian DHP-18. Further evaluation of responsiveness is particularly recommended, and a revised 14-question DHP version is suggested.

Author affiliations
[1]Department of Health and Welfare, Østfold University College, Halden, Norway
[2]Department of Internal Medicine, Østfold Hospital Trust, Grålum, Norway
[3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[4]Department of Endocrinology, Telemark Hospital Trust, Skien, Norway
[5]Morbid Obesity Centre and Section of Endocrinology, Vestfold Hospital Trust, Tønsberg, Norway
[6]Research and Development, Vestfold Hospital Trust, Tønsberg, Norway

Contributors In particular, ØJ, TB and LPJJ contributed to the study design, data analysis and interpretation of data. CG, RBM and DH were the local PIs and were responsible for recruitment at the centers. Furthermore, all authors drafted the work, revised it critically for intellectual content, and approved the final version of the manuscript.

REFERENCES

1. Smith-Palmer J, Bae JP, Boye KS, *et al*. Evaluating health-related quality of life in type 1 diabetes: a systematic literature review of utilities for adults with type 1 diabetes. *Clinicoecon Outcomes Res* 2016;8:559–71.
2. Fayers PM, Machin D. *Quality of life: the assessment, analysis, and reporting of patient-reported outcomes*. 3rd edn. Chichester, West Sussex: John Wiley, 2016.
3. Ware JE, Gandek B, Guyer R, *et al*. Standardizing disease-specific quality of life measures across multiple chronic conditions: development and initial evaluation of the QOL Disease Impact Scale (QDIS®). *Health Qual Life Outcomes* 2016;14:84.
4. Meadows K, Steen N, McColl E, *et al*. The Diabetes Health Profile (DHP): a new instrument for assessing the psychosocial profile of insulin requiring patients--development and psychometric evaluation. *Qual Life Res* 1996;5:242–54.
5. Meadows K. *User Guide. The Diabetes Health Profile (DHP-1 and DHP-18)*, 2012.
6. Lazarus RS, Folkman S. *Stress, appraisal, and coping*. Springer Publishing Company, 1984.
7. Meadows KA, Abrams C, Sandbaek A. Adaptation of the Diabetes Health Profile (DHP-1) for use with patients with Type 2 diabetes mellitus: psychometric evaluation and cross-cultural comparison. *Diabet Med* 2000;17:572–80.
8. Mulhern B, Meadows K. The construct validity and responsiveness of the EQ-5D, SF-6D and Diabetes Health Profile-18 in type 2 diabetes. *Health Qual Life Outcomes* 2014;12:42.
9. Gjersing L, Caplehorn JR, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC Med Res Methodol* 2010;10:13.
10. Wild D, Grove A, Martin M, *et al*. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8:94–104.
11. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
12. Loge JH, Kaasa S, Hjermstad MJ, *et al*. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. *J Clin Epidemiol* 1998;51:1069–76.
13. Jacobson AM, de Groot M, Samson JA. The evaluation of two measures of quality of life in patients with type I and type II diabetes. *Diabetes Care* 1994;17–267–74.
14. Ware JE. The status of health assessment 1994. *Annu Rev Public Health* 1995;16:327–54.
15. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
16. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, N. J: Laurence Erlbaum, 1988: 567.
17. Polonsky WH, Hessler D, Ruedy KJ, *et al*. The impact of continuous glucose monitoring on markers of quality of life in adults with type 1 diabetes: further findings from the DIAMOND randomized clinical trial. *Diabetes Care* 2017;40:736–41.
18. Wojciechowski P, Ryś P, Lipowska A, *et al*. Efficacy and safety comparison of continuous glucose monitoring and self-monitoring of blood glucose in type 1 diabetes: systematic review and meta-analysis. *Pol Arch Med Wewn* 2011;121:333–43.

19. Tabachnick BG, Fidell LS. Using multivariate statistics. *Pearson international edition*. 5th edn. Boston: Pearson/Allyn and Bacon, 2007.
20. Tan ML, Khoo EY, Griva K, *et al*. Diabetes health profile-18 is reliable, valid and sensitive in Singapore. *Ann Acad Med Singapore* 2016;45:383–93.
21. Goddijn P, Bilo H, Meadows K, *et al*. The validity and reliability of the Diabetes Health Profile (DHP) in NIDDM patients referred for insulin therapy. *Qual Life Res* 1996;5:433–42.
22. Jelsness-Jørgensen LP, Moum B, Grimstad T, *et al*. Validity, reliability, and responsiveness of the brief pain inventory in inflammatory bowel disease. *Can J Gastroenterol Hepatol* 2016;2016:5624261:1–10.
23. Stucky BD, Pereira CCA. Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol: measurement in medicine: a practical guide. *Quality of Life Research* 2012;21:371–3.
24. Polit DF. Getting serious about test-retest reliability: a critique of retest research and some recommendations. *Qual Life Res* 2014;23:1713–20.
25. Whitty P, Steen N, Eccles M, *et al*. A new self-completion outcome measure for diabetes: is it responsive to change? *Qual Life Res* 1997;6:407–13.