


BMJ Open New method of measuring subjective well-being: prospective validation study of the 'Daily Experience Sampling Questionnaire' (DESQ) in patients with psoriasis and healthy subjects in Germany

Christine Blome ¹, Natalia Kirsten,¹ Ibrahim Nergiz,² Ulrich Schiffner,² Marina Otten,¹ Matthias Augustin¹

To cite: Blome C, Kirsten N, Nergiz I, *et al.* New method of measuring subjective well-being: prospective validation study of the 'Daily Experience Sampling Questionnaire' (DESQ) in patients with psoriasis and healthy subjects in Germany. *BMJ Open* 2020;**10**:e039227. doi:10.1136/bmjopen-2020-039227

► Prepublication history and additional materials for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-039227>).

Received 08 April 2020

Revised 21 October 2020

Accepted 29 October 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Christine Blome;
c.blome@uke.de

ABSTRACT

Objectives To validate the newly developed Daily Experience Sampling Questionnaire (DESQ) that measures affective subjective well-being (SWB). The DESQ is an end-of-day diary in which respondents retrospectively rate their SWB at six different, randomly determined moments; it is completed over 1 week. The DESQ shall provide an alternative or complementary approach to existing methods of near-time SWB measurement (experience sampling, Day Reconstruction Method). The primary research objective was to determine criterion validity of the DESQ.

Design Prospective, non-interventional study.

Setting Participants were recruited in Hamburg, Germany, at a specialised outpatient clinic (patients) and via different channels (healthy participants).

Participants 101 adults with diagnosed and stable psoriasis (46 women, 55 men); 105 adults without psoriasis (49 women, 56 men).

Primary and secondary outcome measures Participants completed the DESQ for 3 weeks. In weeks 2 and 3, they also performed experience sampling. Criterion validity was determined by weekwise intraclass correlations (ICC) between both methods. Sensitivity to change was determined by the correlation between changes in both methods from weeks 2 to 3. For convergent validity, related concepts such as life satisfaction were measured. Retest reliability was determined using DESQ values of weeks 2 and 3.

Results Criterion validity was excellent (ICC: patients=0.86, 95% CI 0.81 to 0.91; healthy participants=0.86, 95% CI 0.79 to 0.91). Sensitivity to change was $r=0.57$ and $r=0.56$, respectively. Correlations with convergent criteria were mostly significant and higher in constructs more proximal to SWB. The ICC indicating retest reliability was 0.77 in patients (95% CI 0.68 to 0.84) and 0.81 in healthy participants (95% CI 0.73 to 0.86).

Conclusions The DESQ is a valid, reliable and feasible instrument for SWB measurement in people with psoriasis and healthy people. Its approach of end-of-day evaluations of single moments may also lend itself to the

Strengths and limitations of this study

- This validation study included both patients and controls matched for age and gender.
- The experience sampling method was used as a validation criterion, enabling the evaluation of criterion validity. This is not possible in the majority of questionnaire validation studies due to the absence of a gold standard measure.
- As most patients in this study currently had only mild psoriasis, results cannot necessarily be generalised to more severely affected patients.
- More healthy participants had higher education, which limits comparability with the patient group.
- The Daily Experience Sampling Questionnaire and the experience sampling method had to be used in parallel in order to estimate convergence between both methods, which may have resulted in memory effects.

measurement of other highly time-variant constructs such as pain, fatigue or depression.

INTRODUCTION

Subjective well-being (SWB) can be defined as affective SWB, that is, the extent of positive and negative emotions a person experiences, or cognitive SWB, that is, satisfaction with life.¹ This study focused on affective SWB.

SWB is important not only because people strive to experience it, but also because it predicts other positive outcomes such as altruistic behaviour or the experience of meaning.² SWB is also important in the medical context. Patient well-being is increasingly considered an important outcome that should be assessed as an endpoint within clinical trials, but also in clinical practice.³ Arguably, SWB



is of particular importance in chronic conditions where patients experience impairments for a long period of time. One such chronic disease with potentially marked SWB impairments⁴ is psoriasis, which is why we chose to include people with this condition in our study. Psoriasis is a chronic, systemic, immune-mediated disease that leads to flaking and scaling of the skin, often accompanied by joint or nail problems. Its prevalence varies considerably between countries; in Germany, psoriasis affects 2%–3% of adults.⁵ Health-related quality of life (HRQoL) can be impaired to a degree comparable with diabetes or cancer; patients suffer from physical symptoms such as pain and itching, but also from psychosocial effects of stigmatisation, depression and treatment burden.⁶ As underlined in the World Health Organization Global report on psoriasis (2016), psoriasis often has considerable impact on daily life, including social activities and work life, and mental well-being, up to the point of being ‘psychologically devastating’.⁷ The World Health Assembly (WHA) has identified psoriasis as one of the five major non-communicable diseases.⁸

This study also included a group of healthy participants in order to provide a basis for Daily Experience Sampling Questionnaire (DESQ) usage outside the medical field, for example, in positive psychology or the economic sciences, and for DESQ usage in a healthy control group within medical studies.

Patient well-being is mostly evaluated using questionnaires on HRQoL, which is a multidimensional construct including physical, social, emotional and other influences of illness on a patient’s quality of life.⁹ HRQoL instruments therefore focus on illness-related aspects of well-being; they do not assess well-being directly (eg, by asking for positive and negative emotions), but indirectly by asking for circumstances assumed to influence well-being (eg, physical symptoms).¹⁰ However, the same circumstances can have very different effects on SWB, which may impair comparability of HRQoL assessments both intraindividually and interindividually.¹⁰ Therefore, measuring SWB may be an important complement to HRQoL in determining patients’ subjective impairment.

SWB has also been proposed as a measure of health utilities,¹¹ that is, ratio-scaled measures of the utility of a particular health state with 0 representing *as bad as being dead* and 1 representing *best imaginable health*; values below 0 indicate health states worse than being dead. Health utilities are used, among others, to calculate quality-adjusted life years in health-economic evaluations so that both the duration and the quality of a health state are taken into account.

Health utilities are usually determined by eliciting the general population’s preference for the health state in question based on fictional scenarios.¹² This approach has been criticised, among others, for its susceptibility to focusing illusion bias, which means that its focus on health produces overly negative evaluations of the health state in question.¹³ This is why experienced utilities have been proposed as an alternative: measuring SWB of patients

who experience the health state of interest without explicitly referring to health.¹¹ Here, health utilities can be calculated as the average percentage of moments that people in the health state experience positive affect (a measure called U-index).¹¹

There are different methods of measuring SWB. In the Positive and Negative Affect Schedule (PANAS),¹⁴ respondents rate the extent to which they experienced different positive and negative emotions within a defined period of time. This approach risks recall bias due to more extreme episodes and more recent episodes being overrepresented in memory (peak-end rule).^{15 16}

Recall bias is avoided in the experience sampling method (ESM) where people repeatedly rate their momentary SWB at randomly determined moments on a mobile device. This method has been used extensively in psychological research for measuring SWB and other constructs.¹⁷ It captures not only intensity, but also dynamics and temporal patterns.¹⁸

But ESM also has methodological and practical drawbacks. It causes multiple interruptions during a day¹⁹; devices must be provided if participants do not have compatible devices; technical briefing and troubleshooting is often required. Missing values arise because responding is impractical in many situations, such as when taking a shower or going to the movies. In many occupations, using a phone may not be permitted or possible, resulting in the exclusion of potential participants (eg, nurses).

To avoid these drawbacks, the Day Reconstruction Method (DRM) was developed.²⁰ In this end-of-day measure, participants divide the complete day into episodes, note start and end time of each episode, and rate each episode for SWB. There are no interruptions during the day, and the measure can be completed on paper or electronically. High convergence was found between DRM and ESM data,^{19 21} but affect was rated systematically lower in the DRM than in the ESM.²²

Though cheap and easy to administer, the DRM has the drawback of being time consuming,¹⁹ which is why it is used mainly for single-day studies. In addition, respondents need to make a global SWB judgement of each episode, even if their well-being during the episode fluctuated. DRM data can thus be biased by peak-end effects.¹⁹ The latter problem also applies to short versions of the DRM where only parts of the day are evaluated.^{23 24}

In this study, the newly developed DESQ was used. It combines advantages of ESM and DRM by using a once-daily assessment via questionnaires (like DRM) and assessing only a random sample of moments (like ESM). The DESQ is completed each evening over seven days. For each day, it covers six time points (eg, 14:30), which have been determined randomly and which vary between days. Respondents state their recalled activity at that moment (eg, shopping) and rate their recalled SWB on a seven-point smiley scale ranging from *very bad* to *very good* (figure 1). Thereby, they are asked for their global SWB on a single good-bad-dimension instead of rating

"DESQ" – QUESTIONNAIRE ON WELL-BEING

Please try to remember what you did **today, Monday** _____, at the time points listed below and how you felt at the respective moments.

If you do not remember exactly, please give an estimation!

If you were **asleep** at a time point, you do not need to choose an answer.

What did you do today at 6:20 a.m.? (Please note briefly)

How did you feel at 6:20 a.m.?

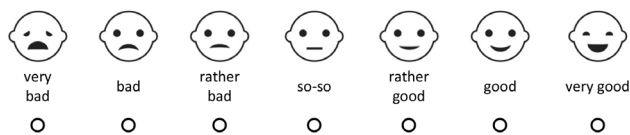


Figure 1 Extract from Daily Experience Sampling Questionnaire (DESQ) (own translation from the German original).

different positive and negative emotions, as usually done in PANAS, ESM, and DRM. The assumption underlying the use of this scale is that the valence of almost every (single) moment can be located on either the good or bad side of neutral, even though sometimes different good and bad emotions may be present at the same time.²⁵ This unidimensionality of SWB in single moments is supported by latent state-trait models of momentary SWB data, showing that ‘situational influences that make people happier are exactly those that make people less unhappy’.²⁶

The DESQ had been developed with both psoriasis patients and healthy people,²⁷ including a survey on response options for the SWB scale and one week completion of the questionnaire followed by cognitive debriefing.

The aim of this study was to validate the DESQ in people with psoriasis and healthy people in order to evaluate it for use in studies both in the medical context and in studies on the general population. The primary research question of this study was whether the DESQ has criterion validity, using ESM as a gold standard.²⁸ ESM was chosen as gold standard of SWB measurement because its real-time approach minimises recall bias. However, its practical limitations motivated the development of the DESQ; if DESQ data converge with ESM data despite of the DESQ’s longer recall period, this suggests that valid data can also be collected with this more practical approach. Additional properties to be tested included convergent validity, retest reliability, sensitivity to change and feasibility.

METHODS

Setting and participants

Participants were at least 18 years old and had sufficient cognitive and German language abilities to complete the measures.

Participants with psoriasis (in the following: patients) had received the diagnosis a year or more ago; physicians judged that severity of psoriasis would probably remain stable during data assessment; treatment had not changed in the previous three months and was expected to remain unchanged during data assessment.

The only inclusion criterion specific to the group of participants without psoriasis (in the following: healthy participants) was not having psoriasis. They were recruited to match the patient group in terms of age group and gender.

Most patients were recruited at a specialised outpatient clinic for psoriasis; healthy participants were recruited via different channels (table 1). All participants lived in the city of Hamburg, Germany, and surroundings.

Procedures

On study inclusion, the ESM software (see below) was installed on the participant’s own mobile phone, if compatible, or a loaned device. Participants tried out a demo version and were given a folder including the paper questionnaires (see below).

Starting the following Sunday, participants completed paper questionnaires every evening for 22 consecutive days (figure 2). These covered sociodemographic and clinical data (day 1), the DESQ (days 2–22), a number of convergent criteria (baseline or once weekly), and feasibility questions (day 22).

In the second and third weeks, participants also completed the ESM on the mobile phone. The reason for starting ESM in week 2 only is that we wanted to determine whether the DESQ responses changed systematically once participants also completed the ESM; this would suggest reactivity,²⁹ in particular due to learning effects.

A compensation of 100 to 150 Euro, depending on ESM data completeness, was paid.

Instruments

Age, gender, education, and employment situation were assessed at baseline.

Comorbidity was assessed with the German Self-Administered Comorbidity Questionnaire (SCQ-D).^{30 31}

Psoriasis characteristics included year of initial diagnosis, whether nails or joints were affected by psoriasis, and whether psoriatic arthritis had been diagnosed. A physician global assessment of current disease severity was obtained on a visual analogue scale broken down into 5 segments labelled 0=no, 1=mild, 2=moderate, 3=sever, 4=very severe psoriasis.

The DESQ should be completed every evening. For each week and person, a total score was calculated as the U-index. This index is the percentage of moments a person feels good. It was proposed by Dolan and Kahneman¹¹ as a ratio-scaled SWB score—a feature that makes it suitable, among others, as a utility measure in health-economic evaluation. This is why Dolan and Kahneman¹¹ recommend the U-index as a health utility measure instead of the commonly used preference-based measures. In this study, the U-index was calculated for

Table 1 Baseline characteristics of study participants

	Patients (n=101)	Healthy participants (n=105)
Age: mean±SD, range	46.6±13.9, 19–75	45.9±14.7, 19–78
Male gender: n (%)	55 (54.5)	56 (53.3)
School education degree: n (%)		
General education (9 years)	17 (16.8)	4 (3.8)
Middle school (10 years)	36 (34.7)	26 (24.8)
Higher education (12 or 13 years)	48 (47.5)	74 (70.5)
Missing	–	1 (1.0)
Job situation: n (%)		
Working (employed or self-employed)	76 (75.2)	75 (71.4)
Working hours per week: mean±SD	36.4±9.2 (1 missing)	33.0±11.7 (4 missings)
Not working: n (%)	25 (24.8)	29 (27.6)
Leave of absence (eg, family leave)	3 (3.0)	2 (1.0)
Trainee or voluntary social year	1 (1.0)	4 (3.8)
Retired or prematurely retired (not due to disease)	10 (9.9)	13 (12.4)
Prematurely retired (due to disease)	3 (3.0)	1 (1.0)
Homemaker	5 (5.0)	2 (1.9)
Student (incl. pupils)	4 (4.0)	4 (3.8)
Unemployed	6 (5.9)	4 (3.8)
Rehabilitation	–	1 (1.0)
Current psoriasis severity (physician global assessment): mean±SD, range	0.8±0.7, 0–3 (4 missings)	n.a.
Years since first diagnosis: mean±SD, range	25.7±14.5, 1–55 (17 missings)	n.a.
Nails affected by psoriasis: n (%)		
Yes, currently	28 (28.8)	n.a.
Yes, formerly	33 (32.7)	n.a.
No, never	38 (37.6)	n.a.
Don't know/ no reply	2 (2.0)	n.a.
Joints affected by psoriasis: n (%)		
Yes, currently	31 (30.7)	n.a.
Yes, formerly	14 (13.9)	n.a.
No, never	47 (46.5)	n.a.
Don't know/no reply	9 (8.9)	n.a.
Diagnosis of psoriatic arthritis: n (%)		
Yes	21 (20.8)	n.a.
No	69 (68.3)	n.a.
Don't know/no reply	11 (10.9)	n.a.
Place of recruitment: n (%)		
Outpatient clinic for psoriasis	98 (97.0)	–

Continued

Table 1 Continued

	Patients (n=101)	Healthy participants (n=105)
Word-of-mouth	2 (2.0)	0 (39.0)
Email newsletter to employees	–	20 (19.0)
Job posting website for students	1 (1.0)	14 (13.3)
Dental clinic	–	13 (12.4)
Short article in a local newspaper	–	7 (6.7)
Dental practice	–	6 (5.7)
Other	–	4 (3.8)
EQ-5D-3L on subjective health: mean±SD, range	0.83±0.16, 0.25–1.00 (1 missing)	0.89±0.14, 0.38–0.1
EQ VAS on subjective health: mean±SD, range	74.4±16.9, 20–100	82.2±14.4, 35–100
Skindex-17 (score on psychosocial impairment): mean±SD, range	3.3±4.5, 0–22 (3 missings)	n.a.
Skindex-17 (score on symptoms): mean±SD, range	2.5±1.9, 0–6 (2 missings)	n.a.
SCQ-D: mean±SD, range	1.6±1.5, 0–7	1.3±1.7, 0–9
NEO-FFI neuroticism score: mean±SD, range	19.5±8.8, 2–39	17.2±8.0, 0–39
SWLS: mean±SD, range	24.1±5.8, 6–35	24.9±6.0, 6–34
PANAS, positive affect subscale: mean±SD, range	3.1±0.6, 1.2–4.3 (3 missings)	3.4±0.6, 1.8–4.5 (7 missings)
PANAS, negative affect subscale: mean±SD, range	1.7±0.5, 1.0–3.3 (5 missings)	1.7±0.6, 1.0–3.7 (2 missings)
ADS-K on depression: mean±SD, range	10.0±7.6, 0–43 (3 missing)	8.0±5.5, 0–26 (2 missing)

ADS-K, Allgemeine Depressionsskala; EQ-5D-3L, EQ-5D questionnaire with three-level response scale; EQ VAS, visual analogue scale within the EQ-5D; n.a., not applicable; NEO-FFI, NEO-Five-Factor-Inventory; PANAS, Positive and Negative Affect Schedule; SCQ-D, German Self-Administered Comorbidity Questionnaire; Skindex-17, Skin Disease-Specific Quality of Life Questionnaire; SWLS, Satisfaction with Life Scale.

each participant by dividing the number of the responses 'rather good', 'good', or 'very good' by the total number of valid responses of the respective person, and multiplying the result by 100.

For the ESM, we used movisensXS for Android (Movisens GmbH, Karlsruhe, Germany). This app sent acoustic alerts at a random time within each 3-hours interval between 05:00 and 23:00, except when participants declared themselves to be asleep using an app function for this purpose. Participants thus received up to six alerts per day. Every alarm was repeated four times after an interval of 1 min each. If participants forgot to declare themselves awake by 10:00 a.m., the app would

Day #	Week 1 (Monday-Sunday)							Week 2 (Monday-Sunday)							Week 3 (Monday-Sunday)							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Socio-demographic data																						
Clinical data on psoriasis (only patients)																						
Self-Administered Comorbidity Questionnaire (SCQ-D)																						
Neuroticism (NEO-FFI subscale)																						
Satisfaction with Life Scale (SWLS)																						
Positive and Negative Affect Schedule (PANAS)																						
Depression questionnaire (ADS-K)																						
EuroQoL questionnaire (EQ-5D)																						
Skin disease-specific quality of life questionnaire (Skindex-17, only patients)																						
Subjective well-being (SWB) in the preceding week (single item)																						
Experience sampling method (ESM) via mobile phone with up to 6 assessments per day																						
Daily Experience Sampling Questionnaire (DESQ)																						
Feasibility questions																						

Figure 2 Schedule of data assessment for each participant.

automatically resume sending alerts. Through the alert, participants were asked to rate their SWB at that moment on a 7-point scale (the same as included in the DESQ, but in vertical orientation; see online supplemental figure 1). For each week of the ESM, the U-index was calculated.

As convergent criteria, we chose constructs assumed to be associated with SWB as they comprised a subjective evaluation of quality of life and/or were related to positive or negative affect. These variables were assessed weekly on Sunday evenings:

- ▶ Subjective health on the present day was measured with EQ-5D questionnaire with three-level response scale (EQ-5D-3L) and EQ VAS.³² The EQ-5D-3L assesses mobility, self-care, usual activities, pain/discomfort, and anxiety/depression with one item each. From these, a health utility score ranging 0–1 was calculated using the German valuation set representing societal values of the general public.³³ In the EQ VAS, participants rate their health on a visual analogue scale from 0 (worst) to 100 (best imaginable).
- ▶ Skin-related HRQoL within the last week was assessed with the Skin Disease-Specific Quality of Life Questionnaire (Skindex-17),^{34 35} from which sum scores on psychosocial impairment and symptoms were calculated.
- ▶ Depression in the last week was measured with the 15-item Allgemeine Depressionsskala (ADS-K),³⁶ a German short version of the Center for Epidemiologic Studies Depression Scale³⁷; the sum score was calculated.
- ▶ The 5-item Satisfaction with Life Scale (SWLS)^{38 39} does not refer to a specific time period; a global score was calculated.
- ▶ The PANAS asked for the extent of ten positive and ten negative emotions within the last week,¹⁴ resulting in two sum scores.
- ▶ SWB in the preceding week was assessed with a global rating, using the same 7-point rating scale as in the DESQ.
- ▶ Neuroticism—the tendency to experience negative emotions—was assessed with the 60-item NEO-FFI (NEO-Fünf-Faktoren-Inventar; NEO-Five-Factor-Inventory), of which 12 items were used to

calculate the neuroticism score.⁴⁰ It was assessed at baseline only as neuroticism is a personality trait that is highly stable over time.⁴¹

Feasibility questions with 4-point response scales covered: comprehensibility of instructions, ease of retrospectively stating one's activities and SWB, and feasibility of completing the questionnaire every evening. Participants stated their preference for DESQ or ESM, or chose undecided.

Statistics

Data were analysed with SPSS V.22 (IBM, Armonk, New York).

For criterion validity, agreement between DESQ and ESM in weeks 2 and 3, respectively, was determined through two-way mixed, average score, absolute agreement intraclass correlation coefficients (ICC).

The sample size of 100 participants per group was sufficient to determine the 95% CI of pairwise differences between DESQ and ESM in a Bland-Altman plot with a precision of ± 0.34 SD in each group.

Sensitivity to change was determined as Spearman correlation of change scores in DESQ and ESM from week 2 to week 3. For convergent validity, Spearman correlations between DESQ and convergent criteria were calculated. For the EQ-5D-3L, ICCs were also calculated, as it likewise represents a utility score ranging from 0 to 1. Therefore, its absolute agreement with the DESQ was of interest. For retest reliability, ICC between DESQ values of weeks 2 and 3 was calculated. As SWB is assumed to fluctuate, convergence between weeks 2 and 3 values should be interpreted as resulting from an individual tendency to experience higher or lower levels of SWB. In addition, we determined the ICC for week 1 and 3 (ie, longer retest period) and for weeks 1 and 2 (ie, 1 week with and one without concurrent ESM assessment).

RESULTS

Participants

Out of 116 patients who gave informed consent, 101 were analysed; of 113 healthy participants, 105 were analysed

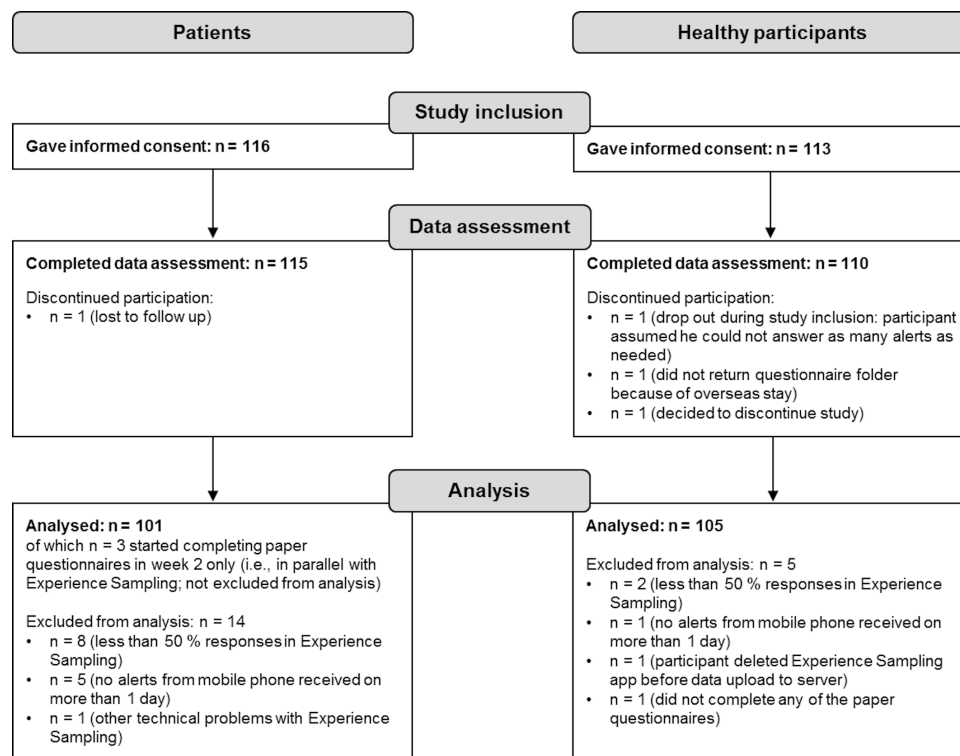


Figure 3 Flow chart on study inclusion. N, number of participants.

(figure 3). The most common reason for exclusion was acting on less than 50% of ESM alerts.

Due to matching, both groups were similar in age and gender (table 1). However, more healthy participants than patients had higher education.

Average psoriasis severity was low at 0.8 (possible range 0–4), which was a result of the inclusion criterion ‘no change in severity and treatment to be expected’: at the specialised outpatient clinic, treatment (often systemic) will usually be adapted until psoriasis has improved. Therefore, most patients had a history of severe psoriasis but low current disease activity. Accordingly, skin-specific HRQoL was quite good, with 19 patients (18.8%) attaining the best possible value in both Skindex subscales.

Subjective well-being

The distribution of SWB responses was highly skewed, the most common response being *good* in both DESQ and ESM (39.8%–43.4% of valid responses; figure 4). The rating *very bad* was chosen least often at 0.1%–0.2% of moments.

The average DESQ U-Index was similar across the three weeks in both patients (79.1, 78.4, 78.1) and healthy participants (82.6, 82.8, 81.8; table 2). The group difference was not significant (t test for independent samples: week 1, $p=0.174$; week 2, $p=0.175$; week 3, $p=0.309$). During the same time, the average ESM U-index increased slightly from 78.8 to 80.7 in patients and from 79.8 to 81.2 in healthy participants.

Testing for systematic changes in the DESQ from week 1 to week 2 (when the ESM was added), the paired samples t-test was not significant in either patients ($p=0.698$) or

healthy participants ($p=0.914$), neither was the Wilcoxon signed-rank test for related samples on changes in median (patients: $p=0.974$; healthy participants: $p=0.313$).

Psychometric properties

For criterion validity of the DESQ (table 2) at week 2, an ICC of 0.86 was found in both groups. At week 3, ICC was 0.85 in patients and 0.90 in healthy participants. With all ICC values higher than 0.75, this can be considered excellent agreement.⁴² Bland-Altman plots (figure 5) visualise the agreement, plotting differences between DESQ and ESM against the mean of both measurements. The lines for upper and lower limit of agreement show where about 95% of the differences between DESQ and ESM lie, which in week 2 was between -26 and $+23$ points in patients and between -18 and $+24$ points in healthy participants. Visually, these differences (Y axis) do not appear to differ systematically by level of SWB (X-axis).

For retest reliability of the DESQ (table 2) in week 2 to 3, ICC was 0.77 in patients and 0.81 in healthy participants, indicating excellent concordance.⁴² For ESM, values were similar with ICC=0.75 and 0.78. For weeks 1 and 2 (ie, 1 week with and one without concurrent ESM assessment) and weeks 1–3 (ie, longer retest period), ICCs were lower.

Sensitivity to change of the DESQ (table 2) was highly significant with a change correlation of $r=0.57$ in patients and $r=0.56$ in healthy participants.

All correlations determined for convergent validity of the DESQ (table 3) were significant, except for the Skindex psychosocial subscale in week 3 ($p=0.052$). The level of correlations differed between weeks, but without a clear pattern of increase or decrease. In both groups,

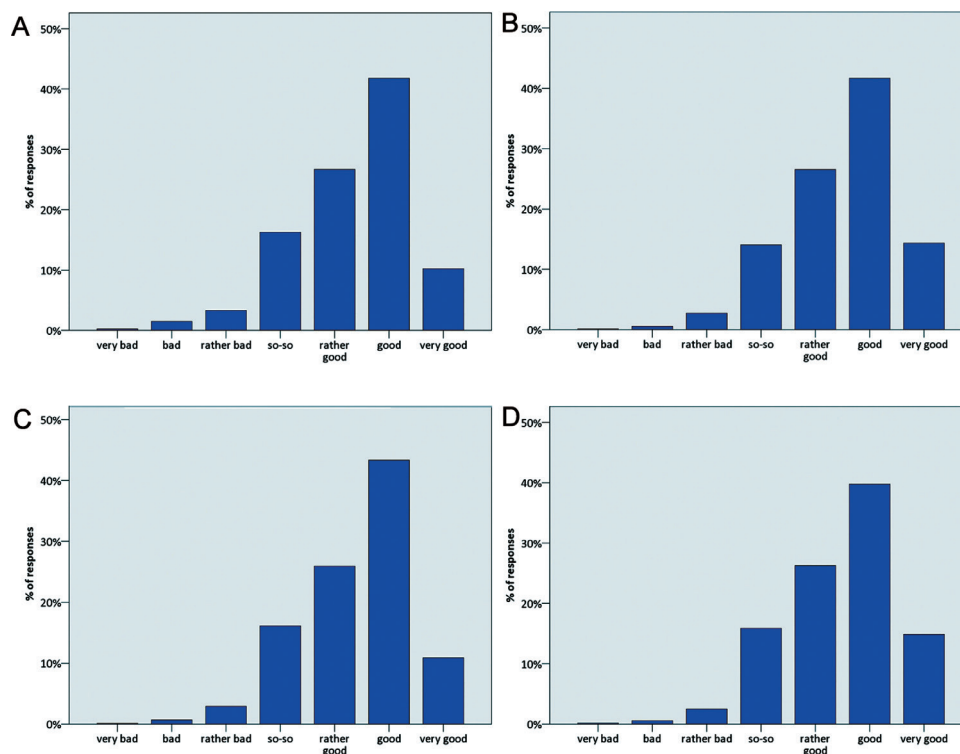


Figure 4 Distribution of single responses to the subjective well-being item. (A) Patient (n=10 495) and (B) healthy participant responses (n=10 968) within the Daily Experience Sampling Questionnaire (days 2–22), (C) patient (n=6875) and (D) healthy participant (n=7094) responses within the experience sampling method (days 15–22).

the highest correlations were found for the SWB single item (up to $r=0.75$) and the ADS-K depression score (up to $r=-0.66$). Correlations with life satisfaction, neuroticism and subjective health were weak to moderate.

Agreement between DESQ and EQ-5D measures was around ICC=0.4 in patients (threshold between poor and fair agreement⁴²) and even lower in two out of 3 weeks in healthy participants.

Table 2 Distribution and reliability of Daily Experience Sampling Questionnaire (DESQ) and experience sampling method (ESM); criterion validity and sensitivity to change of DESQ

	DESQ mean±SD	ESM mean±SD (median)	DESQ mean±SD	ESM mean±SD (median)
	Patients (n=101)		Healthy participants (n=105)	
SWB week 1	79.1±21.3 (84.8)	n.a.	82.6±14.2 (85.3)	n.a.
SWB week 2	78.4±25.1 (86.5)	78.8±21.8 (86.5)	82.8±20.7 (90.3)	79.8±20.2 (84.6)
SWB week 3	78.1±28.4 (90.6)	80.7±25.4 (91.7)	81.8±23.2 (92.1)	81.2±22.5 (89.2)
Retest reliability: ICC between week 2 and week 3 (95% CI)	0.77 (0.68 to 0.84)	0.75 (0.65 to 0.83)	0.81 (0.73 to 0.86)	0.78 (0.69 to 0.85)
Retest reliability: ICC between week 1 and week 2 (95% CI)	0.68 (0.56 to 0.77)	n.a.	0.45 (0.28 to 0.59)	n.a.
Retest reliability: ICC between week 1 and week 3 (95% CI)	0.59 (0.44 to 0.70)	n.a.	0.31 (0.12 to 0.47)	n.a.
Criterion validity: ICC between DESQ and ESM (95% CI), week 2	0.86 (0.81 to 0.91)		0.86 (0.79 to 0.91)	
Criterion validity: ICC between DESQ and ESM (95% CI), week 3	0.85 (0.79 to 0.90)		0.90 (0.85 to 0.93)	
Sensitivity to change: Correlation of changes in DESQ and ESM from week 2–3	r=0.57 (p<0.001)		r=0.56 (p<0.001)	

ICC, intraclass correlation coefficient; n.a., not applicable (as ESM started in week two only); r, Spearman correlation coefficient; SWB, subjective well-being.

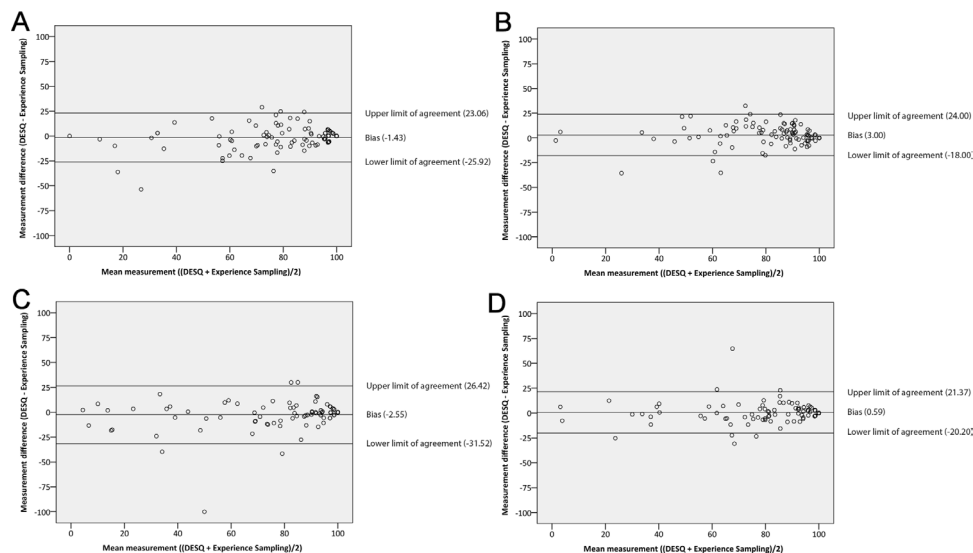


Figure 5 Bland-Altman plots showing the level of agreement between the Daily Experience Sampling Questionnaire (DESQ) and experience sampling method (ESM; U-indexes). (A) Patient and (B) healthy participant group in week 2, (C) patient and (D) healthy participant group in week 3.

Most patients (99.0%) and healthy participants (99.1%) rated the DESQ instruction as (rather) easy to understand (table 4). The DESQ activity free text items were rated mostly/(almost) always easy to answer by 92.0% of patients and 95.3% of healthy participants. Almost as many respondents (90.1%; 91.4%) regarded the DESQ SWB items as (mostly) easy to answer. Most found daily completion of the DESQ hardly/not at all bothersome (96.1%; 96.2%). Overall, 41.6% of patients and 51.4% of healthy participants preferred DESQ over ESM; 34.7%

and 29.5% preferred ESM (undecided: 21.8% and 17.1%).

The most common reason for missing responses to the DESQ was that participants were asleep, either as documented in the activity free text or as assumed by us because both activity and SWB were missing only for the first time point of the day (table 5). The remaining non-entries were considered genuine missing values.

In ESM, the most common reason for missing values was that respondents had declared being asleep with the

Table 3 Convergent validity of the Daily Experience Sampling Questionnaire (DESQ) U-index (weeks 1–3)

	Patients (n=101)			Healthy participants (n=105)		
	Week 1	Week 2	Week 3	Week 1	Week 2	Week 3
EQ-5D-3L on subjective health: ICC (95% CI)	0.38 (0.20 to 0.53)	0.40 (0.22 to 0.55)	0.40 (0.22 to 0.55)	0.23 (0.05 to 0.41)	0.40 (0.22 to 0.55)	0.33 (0.14 to 0.49)
EQ-5D-3L on subjective health: r (p)	0.43 (<0.001)	0.38 (<0.001)	0.34 (<0.001)	0.32 (0.001)	0.44 (<0.001)	0.31 (0.001)
EQ VAS on subjective health: r (p)	0.55 (<0.001)	0.51 (<0.001)	0.44 (<0.001)	0.29 (0.004)	0.33 (<0.001)	0.39 (<0.001)
Skindex-17 subscale <i>psychosocial impairments</i> : r (p)	-0.42 (<0.001)	-0.22 (0.0035)	-0.25 (0.017)	n.a.	n.a.	n.a.
Skindex-17 subscale <i>symptoms</i> : r (p)	-0.37 (<0.001)	-0.19 (0.052)	-0.22 (0.032)	n.a.	n.a.	n.a.
PANAS subscale on positive emotions: r (p)	0.56 (<0.001)	0.60 (<0.001)	0.42 (<0.001)	0.46 (<0.001)	0.36 (<0.001)	0.54 (<0.001)
PANAS subscale on negative emotions: r (p)	-0.52 (<0.001)	-0.53 (<0.001)	-0.39 (<0.001)	-0.41 (<0.001)	-0.46 (<0.001)	-0.36 (<0.001)
Single item on SWB in the preceding week: r (p)	0.72 (<0.001)	0.72 (<0.001)	0.69 (<0.001)	0.65 (<0.001)	0.65 (<0.001)	0.75 (<0.001)
SWLS on satisfaction with life: r (p)	0.28 (0.004)	0.51 (<0.001)	0.35 (<0.001)	0.30 (0.002)	0.40 (<0.001)	0.35 (<0.001)
ADS-K on depression: r (p)	-0.66 (<0.001)	-0.64 (<0.001)	-0.52 (<0.001)	-0.56 (<0.001)	-0.56 (<0.001)	-0.63 (<0.001)
NEO-FFI subscale on neuroticism: r (p)	-0.43 (<0.001)	-0.48 (<0.001)	-0.36 (<0.001)	-0.29 (0.003)	-0.32 (0.01)	-0.23 (0.021)

ADS-K, Allgemeine Depressionsskala; EQ-5D-3L, EQ-5D questionnaire with three-level response scale; ICC, intraclass correlation coefficient; n.a., not applicable (Skindex-17 was not assessed in healthy participants); p, level of significance; PANAS, Positive and Negative Affect Schedule; r, Spearman correlation coefficient; Skindex-17, skin disease-specific quality of life questionnaire; SWB, subjective well-being; SWLS, Satisfaction with Life Scale; EQ VAS, visual analogue scale within the EQ-5D.

Table 4 Responses to feasibility questions on the DESQ

	Patients (n=101)		Healthy participants (n=105)	
	N	%	N	%
'The instruction on how to complete the questionnaire was...				
...easy to understand'	93	92.1	97	92.4
...rather easy to understand'	7	6.9	7	6.7
...rather difficult to understand'	1	1.0	0	99.0
...difficult to understand'	0	0.0	0	0
No response	0	0	1	1.0
'The question on what you did at the single moments was...				
...(almost) always easy to answer'	46	45.5	49	46.7
...mostly easy to answer'	47	46.5	51	48.6
...mostly difficult to answer'	8	7.9	4	3.8
...(almost) always difficult to answer'	0	0	0	99.0
No response	0	0	1	1.0
'The question on how you felt at the single moments was...				
...(almost) always easy to answer'	48	47.5	46	43.8
...mostly easy to answer'	43	42.6	50	47.6
...mostly difficult to answer'	9	8.9	8	7.6
...(almost) always difficult to answer'	1	1.0	0	99.0
No response	0	0	1	1.0
'Daily completion of the DESQ questionnaire was...				
...well feasible in daily life/ not bothersome'	54	53.5	69	65.7
...rather feasible in daily life/ hardly bothersome'	43	42.6	32	30.5
...rather not feasible in daily life/ rather bothersome'	2	2.0	2	1.9
...not feasible in daily life/ bothersome'	1	1.0	1	1.0
No response	1	1.0	1	1.0
'Which method would you prefer as a study participant?'				
'Completing the smiley questionnaire (DESQ) over a week'	42	41.6	54	51.4
'Answering the smiley question at the mobile phone over a week'	35	34.7	31	29.5
'Both equally preferred or not preferred'	22	21.8	18	17.1
No response	2	2.0	2	1.9

DESQ, Daily Experience Sampling Questionnaire.

respective app function. Further reasons included alerts being ignored or dismissed, late responses, incomplete responses (eg, not pressing the check mark after choosing a response), switching the phone off, and technical problems. The overall rate of missing values was slightly higher

Table 5 Missing values in subjective well-being (SWB) variable by method (Daily Experience Sampling Questionnaire (DESQ) vs experience sampling method (ESM)) and reason for missingness

	Week 1			Week 2			Week 3		
	M	SD	%*	M	SD	%*	M	SD	%*
Patients (n=101)									
DESQ									
Asleep, according to free-text entry	5.8	2.8	13.7	5.9	3.0	14.1	5.9	3.2	14.1
Apparently asleep (first activity and first SWB rating missing)	0.5	1.5	1.2	0.6	1.7	1.3	0.7	1.9	1.6
No entry (though not documented as sleeping or apparently asleep)	0.7	1.5	1.6	1.0	1.7	2.3	1.1	3.7	2.7
Total	7.0	2.6	16.6	7.5	3.1	17.7	7.7	4.5	18.3
ESM									
Alert ignored	n.a.	n.a.	n.a.	3.6	3.3	8.6	3.2	3.3	7.7
Alert dismissed	n.a.	n.a.	n.a.	0.2	0.5	0.5	0.1	0.4	0.2
Response too late (>10 min after trigger)	n.a.	n.a.	n.a.	0.1	0.8	0.3	0.2	0.8	0.4
Response incomplete	n.a.	n.a.	n.a.	0.0	0.2	0.1	0.0	0.1	0.1
App or phone inactivated by participant	n.a.	n.a.	n.a.	4.0	2.8	9.5	4.4	2.8	10.5
No alert received due to technical problems	n.a.	n.a.	n.a.	0.1	0.5	0.1	0.0	0.0	0.0
Total	n.a.	n.a.	n.a.	8.0	4.4	19.1	7.9	4.4	18.9
Healthy participants (n=105)									
DESQ									
Asleep, according to free text entry	6.3	2.8	15.1	6.6	2.8	15.7	6.2	2.6	14.8
Apparently asleep (first activity and first SWB rating missing)	0.2	0.8	0.4	0.2	0.9	0.5	0.3	1.2	0.7
No entry (though not documented as sleeping or apparently asleep)	0.4	0.8	1.0	0.6	1.2	1.4	0.7	1.4	1.6
Total	6.9	2.8	16.5	7.4	2.8	17.7	7.2	2.6	17.1
ESM									
Alert ignored	n.a.	n.a.	n.a.	3.5	3.9	8.3	4.0	4.2	9.6
Alert dismissed	n.a.	n.a.	n.a.	0.2	0.5	0.4	0.1	0.3	0.2
Response too late (>10 min after trigger)	n.a.	n.a.	n.a.	0.2	1.3	0.4	0.3	1.8	0.7
Response incomplete	n.a.	n.a.	n.a.	0.0	0.1	0.0	0.0	0.0	0.0

Continued



Table 5 Continued

	Week 1			Week 2			Week 3		
	M	SD	%*	M	SD	%*	M	SD	%*
App or phone inactivated by participant	n.a.	n.a.	n.a.	3.7	2.6	8.9	4.2	2.5	10.1
No alert received due to technical problems	n.a.	n.a.	n.a.	0.1	0.8	0.3	0.2	1.0	0.4
Total	n.a.	n.a.	n.a.	7.7	5.2	18.3	8.8	5.6	21.0

*Percentage of all potential 42 responses within a week (six times per day over 1 week).
M, arithmetic mean; n.a., not applicable (as ESM started in week 2 only).;

than in the DESQ. However, the rate of genuine missing values cannot be determined with certainty, as it was unclear how many missing responses were in fact due to participants being asleep—especially in the ESM. Participants may have forgotten to deactivate the ESM app when going to sleep, then ignoring or dismissing incoming alerts. On the other hand, they may have declared being asleep or switched off the phone during the day in order not to be disturbed. In addition, participants may have forgotten to turn off the sleep function immediately after waking up. Because of this uncertainty, missing values were not replaced in this analysis.

DISCUSSION

Available methods for measuring SWB have limitations, which is why the DESQ was developed. It is completed only once daily on paper so that respondents will not be interrupted multiple times a day and do not need to use a mobile device (both of which are advantages over the ESM). By evaluating only a random sample of moments for SWB instead of the whole day, it can be completed quickly and does not require global judgements on periods of time (both are advantages over the DRM).

As an end-of-day evaluation may still be susceptible to (short-term) recall bias, we tested the DESQ against the real-time ESM as a gold standard in this study. Convergence between both methods was excellent, implying that the DESQ is of similar accuracy as the ESM. Both methods can thus be used interchangeably in future studies, for example by offering patients a choice of their preferred method, or by using DESQ instead of ESM in those participants who do not have a (compatible) smartphone.

Our experience from this study is that the ESM required more personnel and greater expense than the DESQ, including purchase and configuration of loaned devices, participant training using a demo version, and technical troubleshooting at short notice. In addition, there were fewer missing values in the DESQ than in the ESM. We cannot exclude that this might have been due to some respondents backfilling the diary for previous days. However, this obviously did not pose a major problem

as criterion validity was high (ie, concordance with the real-time assessment was excellent). Possible backfilling should also be kept in mind when interpreting the finding that daily completion of the DESQ was not or hardly bothersome to most participants: this statement will only hold true for the frequency with which they actually completed the DESQ.

It is important to note that we examined convergence of the weekwise U-indexes at group level. Thus, when using the DESQ for diagnosis at the individual level and/or for single days, it may diverge from real-time ESM assessment. This is underlined by the finding that in a few participants, there was a considerable difference between DESQ and ESM U-indexes (figure 5).

As a secondary research question, we assessed, and mostly confirmed, additional psychometric properties, including retest reliability, convergent validity, and sensitivity to change. Retest reliability was high even though SWB is assumed to vary with time and thus to have differed between the 2 weeks we used for test and retest. Reliability may therefore be underestimated in this study. Convergent validity was confirmed for most criteria. However, the DESQ correlated only weakly with skin-related HRQoL as measured with the Skindex-17, maybe because this instrument has not been developed specifically for people with psoriasis, or due to floor effects, as most patients currently had only mild psoriasis. The latter fact is also a limitation of this study, as the results cannot necessarily be generalised to more severely affected patients. Low disease severity may also be the reason for the finding that patients had only slightly, and not significantly, lower SWB than healthy participants. For both groups alike, we thereby confirmed the finding that most people feel good most of the time.⁴³ Convergence of the DESQ with the EQ-5D utility index was considerably lower than with the ESM, which underlines the very different conceptualisations of health utility as experienced utility versus decision utility,^{11 13} but may also result from the different time frame of the measures (today in the EQ-5D vs the previous week in the DESQ).

Due to matching, age and gender were similar in both groups, but more healthy participants had higher education, which limits comparability with the patient group. As higher education was found to be associated with better affective SWB,^{44 45} it is even more surprising that in our study, patients reported only slightly, and non-significantly, lower SWB than healthy participants. Another limitation of this study is that DESQ and ESM had to be used in parallel in order to estimate convergence between both methods, which may have resulted in memory effects: evaluating their SWB during the day may have improved SWB recall in the evening. However, different time points were used in both methods, and we found no change in the DESQ when ESM was added at week 2, which speaks against a systematic effect of the ESM on DESQ responses.

Future research should assess whether the DESQ is sensitive to changes with regard also to clinical improvements,

for example, in response to different medical treatments. This would support feasibility of the DESQ as an additional patient-reported outcome in clinical trials. Evaluating this characteristic was not possible in the current study where stable disease severity was an inclusion criterion in order to be able to estimate retest reliability. For use of the DESQ in patients with diseases other than psoriasis, additional psychometric evaluation is needed, not least because the pattern and extent of SWB impairment will likely differ depending on the indication.

The principle of end-of-day evaluations of single moments also lends itself to the measurement of other highly time-variant constructs where it could be tested as an alternative to ESM and DRM. DRM has been used to measure pain, fatigue and depression in older adults.⁴⁶ A version of the DESQ that is adapted to, for example, pain measurement would retrospectively assess this symptom at randomly determined time points of the respective day; this might be an even less time-consuming way to estimate the time people spend in pain.

Finally, some participants told us about benefits from completing the DESQ when they returned lend devices and questionnaire folders. For example, one patient used her DESQ responses as ‘proof’ of her being impaired in everyday life when discussing with her clinician. Another participant reported being surprised by finding that she had rated her SWB when walking the dog as good; she had always considered this activity an unpleasant duty but now appraises it as a recreational activity. These anecdotes suggest that the DESQ might also be useful as a tool for reflecting and supporting one’s own well-being.

CONCLUSIONS

We conclude that the DESQ is reliable, valid and sensitive to change in both persons with and without psoriasis. Criterion validity with regard to ESM was confirmed, with even the lower limit of the 95% CI of the ICC lying above the threshold for excellent convergence. The DESQ was well accepted by most participants. It can thus be recommended as a time-saving and resource-saving alternative to the ESM. In future SWB studies, participants could be offered to choose between the paper-based DESQ and the ESM according to their preferences and/or the availability of their own, compatible smartphone, as both methods generated sufficiently similar data on group level. This would make it possible to also include participants who have limited access to their smartphones during the day (eg, shop assistants). In studies with limited resources, only the DESQ could be used to measure SWB so that loaned devices, technical support and in-person training on use of the ESM app are not necessary.

Author affiliations

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

²Center for Dental and Oral Medicine, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany

Acknowledgements The authors gratefully acknowledge the student assistants Alina Bruhns, Joris Lehmann, Christopher Neht and Ruth Tiede for their help in conducting this study, and Arick Trettel for his support in patient recruitment.

Contributors CB conceptualised the study, coordinated the data collection, conducted the data analysis and drafted the manuscript. NK and MA contributed to study design, data collection in the clinical population and writing the manuscript. IN and US contributed to data collection in the healthy population and writing the manuscript. MO contributed to data analysis and interpretation and to writing the manuscript. All authors read and approved the final version.

Funding This study was supported by a grant from the German Research Foundation (DFG; BL 1350/2-1).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval All procedures were in accordance with ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments. The Ethics Committee of the Hamburg Chamber of Physicians provided ethical approval (reference PV5227). Study personnel informed potential participants verbally about the study and handed out a patient information sheet. This information included procedures, data to be collected, data protection and pseudonymisation, and voluntary participation with the option to quit participation at any time. If they decided to take part, participants gave written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Christine Blome <http://orcid.org/0000-0002-1163-1639>

REFERENCES

- Diener E, Heintzelman SJ, Kushlev K, *et al*. Findings all psychologists should know from the new science on subjective well-being. *Canadian Psychology/Psychologie Canadienne* 2017;58:87–104.
- King LA. Interventions for enhancing subjective well-being: can we make people happier and should we? In: Eid M, Larsen RJ, eds. *The science of subjective well-being*. New York: The Guilford Press, 2008: 431–48.
- Black N, Burke L, Forrest CB, *et al*. Patient-reported outcomes: pathways to better health, better services, and better societies. *Qual Life Res* 2016;25:1103–12.
- Schuster B, Ziehfreund S, Albrecht H, *et al*. Happiness in dermatology: a holistic evaluation of the mental burden of skin diseases. *J Eur Acad Dermatol Venereol* 2020;34:1331–9.
- Michalek IM, Loring B, John SM. A systematic review of worldwide epidemiology of psoriasis. *J Eur Acad Dermatol Venereol* 2017;31:205–12.
- Raho G, Koleva DM, Garattini L, *et al*. The burden of moderate to severe psoriasis: an overview. *Pharmacoeconomics* 2012;30:1005–13.
- World Health Organization. Global report on psoriasis, 2016. Available: http://apps.who.int/iris/bitstream/10665/204417/1/9789241565189_eng.pdf [Accessed 19 Oct 2020].

- 8 World Health Assembly. Resolution WHA67.9, 2014. Available: http://apps.who.int/gb/ebwha/pdf_files/wha67/a67_r9-en.pdf [Accessed 19 Oct 2020].
- 9 Bakas T, McLennon SM, Carpenter JS, *et al.* Systematic review of health-related quality of life models. *Health Qual Life Outcomes* 2012;10:134.
- 10 Blome C, Augustin M. Measuring change in quality of life: bias in prospective and retrospective evaluation. *Value Health* 2015;18:110–5.
- 11 Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008;118:215–34.
- 12 Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culver AJ, Newhouse JP, eds. *Handbook of health economics*. Amsterdam, New York: Elsevier, 2000: Vol. 1B. 1723–60.
- 13 Kahneman D, Sugden R. Experienced utility as a standard of policy evaluation. *Environ Resource Econ* 2005;32:161–81.
- 14 Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 1988;54:1063–70.
- 15 Fredrickson BL, Kahneman D. Duration neglect in retrospective evaluations of affective episodes. *J Pers Soc Psychol* 1993;65:45–55.
- 16 Geng X, Chen Z, Lam W, *et al.* Hedonic evaluation over short and long retention intervals: the mechanism of the Peak-End rule. *J Behav Decis Mak* 2013;26:225–36.
- 17 Scollon CN, Kim-Prieto C, Scollon CN. Experience sampling: promises and pitfalls, strengths and weaknesses. *J Happiness Stud* 2003;4:5–34.
- 18 Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. *Psychol Bull* 2015;141:901–30.
- 19 Diener E, Tay L. Review of the day reconstruction method (DRM). *Soc Indic Res* 2014;116:255–70.
- 20 Kahneman D, Krueger AB, Schkade DA, *et al.* A survey method for characterizing daily life experience: the day reconstruction method. *Science* 2004;306:1776–80.
- 21 Han W, Feng X, Zhang M, *et al.* Mood states and everyday creativity: employing an experience sampling method and a day reconstruction method. *Front Psychol* 2019;10:1698.
- 22 Tweten C, Anusic I, Lucas RE, *et al.* A direct comparison of the experience sampling method (ESM) and the day reconstruction method (DRM). poster presented at: Society for personality and social psychology annual convention, San Diego, CA, 2016.
- 23 Anusic I, Lucas RE, Donnellan MB. The validity of the day reconstruction method in the German socio-economic panel study. *Soc Indic Res* 2017;130:213–32.
- 24 Mellor-Marsá B, Miret M, Abad FJ, *et al.* Measurement invariance of the day reconstruction method: results from the courage in Europe project. *J Happiness Stud* 2016;17:1769–87.
- 25 Kahneman D. Objective happiness. In: Kahneman D, Diener E, Schwarz N, eds. *Well-being: foundations of hedonic psychology*. New York: Russell Sage Foundation, 2003.
- 26 Eid M. Measuring the immeasurable: psychometric modeling of subjective well-being data. In: Eid M, Larsen RJ, eds. *The science of subjective well-being*. New York: The Guilford Press, 2008: 141–67.
- 27 Blome C. The daily experience sampling questionnaire (DESQ): a new approach to measuring subjective well-being research poster. *Qual Life Res* 2015;24:82–3.
- 28 Carstensen LL, Turan B, Scheibe S, *et al.* Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychol Aging* 2011;26:21–33.
- 29 Barta W, Tennen H, Litt M. Measurement reactivity in diary research. In: Mehl MR, Conner TS, eds. *Handbook of research methods for studying daily life*. New York: The Guilford Press, 2012.
- 30 Sangha O, Stucki G, Liang MH, *et al.* The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003;49:156–63.
- 31 Streibelt M, Schmidt C, Brünger M, *et al.* Komorbidität im Patientenurteil - geht das? Validität eines Instruments zur Selbsteinschätzung der Komorbidität (SCQ-D). *Orthopäde* 2012;41:303–10.
- 32 EuroQol Group. EuroQol- a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- 33 Greiner W, Claes C, Busschbach JJV, *et al.* Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 2005;6:124–30.
- 34 Augustin M, Wenninger K, Amon U, *et al.* German adaptation of the Skindex-29 questionnaire on quality of life in dermatology: validation and clinical results. *Dermatology* 2004;209:14–20.
- 35 Chren M-M. The Skindex instruments to measure the effects of skin disease on quality of life. *Dermatol Clin* 2012;30:231–6.
- 36 Hautzinger M, Bailer M, Hofmeister D, *et al.* *Allgemeine depressionsskala (ADS)*. 2nd edn. Göttingen: Hogrefe Verlag, 2012.
- 37 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *APPL Psychol Meas* 1977;1:385–401.
- 38 Diener E, Emmons RA, Larsen RJ, *et al.* The satisfaction with life scale. *J Pers Assess* 1985;49:71–5.
- 39 Emerson SD, Guhn M, Gadermann AM. Measurement invariance of the satisfaction with life scale: reviewing three decades of research. *Qual Life Res* 2017;26:2251–64.
- 40 Borkenau P, Ostendorf F. *NEO-FFI – NEO-Fünf-Faktoren-Inventar nAch Costa und McCrae: Handanweisung*. 2nd edn. Göttingen: Hogrefe Verlag, 2008.
- 41 Widiger TA, Oltmanns JR. Neuroticism is a fundamental domain of personality with enormous public health implications. *World Psychiatry* 2017;16:144–5.
- 42 Guidelines CDV. Criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Asses* 1994;6:284–90.
- 43 Diener E, Kanazawa S, Suh EM, *et al.* Why people are in a generally good mood. *Pers Soc Psychol Rev* 2015;19:235–56.
- 44 Nikolaev B. Does higher education increase hedonic and Eudaimonic happiness? *J Happiness Stud* 2018;19:483–504.
- 45 Cuiñado J, de Gracia FP. Does education affect happiness? Evidence for Spain. *Soc Indic Res* 2012;108:185–96.
- 46 M Smith D. Using the day reconstruction method to quantify time spent suffering among older adults with chronic pain. *Stat Trans New Series* 2015;16:429–40.