



# Validity of the Patient Experiences and Satisfaction with Medications (PESaM) Questionnaire

Merel L. Kimman<sup>1</sup> · Marlies S. Wijsenbeek<sup>2</sup> · Sander M. J. van Kuijk<sup>1</sup> · Kioa L. Wijnsma<sup>3</sup> · Nicole C. A. J. van de Kar<sup>3</sup> · Marjolein Storm<sup>4</sup> · Xana van Jaarsveld<sup>5</sup> · Carmen D. Dirksen<sup>6</sup> · PESaM Collaborating Group

© The Author(s) 2018

## Abstract

**Background** This study assessed the validity and reliability of the generic module of the recently developed Patient Experiences and Satisfaction with Medications (PESaM) questionnaire in a sample of patients in the Netherlands.

**Methods** The generic module of the PESaM questionnaire consists of 18 items related to the domains effectiveness, side effects and ease of use of medications. It assesses patients' experiences regarding the impact of the medication on daily life, health and satisfaction. In 2017, the PESaM questionnaire was sent out to idiopathic pulmonary fibrosis patients using pirfenidone or nintedanib, atypical haemolytic uraemic syndrome patients receiving eculizumab and patients using tacrolimus after kidney transplantation. Mean scores for each domain were calculated applying a scoring algorithm. Construct validity and reliability were assessed using recommended methods.

**Results** 188 participants completed the generic module, of whom 48% used pirfenidone, 36% nintedanib, 11% tacrolimus and 5% eculizumab. The generic module has good structural properties. Internal consistency values of the domains were satisfactory (i.e. Cronbach's coefficient alpha above 0.7). Confirmatory factor analysis provided further evidence for construct validity, with good convergent and discriminant validity. The PESaM questionnaire also showed different scores for patients using different medications, in line with expectations, and was therefore able to differentiate between patient groups. Test-retest reliability of the items and domains were rated as moderate to fair (i.e. intraclass coefficients ranged between 0.18 and 0.76).

**Conclusions** The PESaM questionnaire is a unique patient-reported outcome measure evaluating patient experiences and satisfaction with medications. It has been developed in conjunction with patients, ensuring coverage of domains and issues relevant from the patient's perspective. This study has shown promising validity of the generic module of the PESaM questionnaire. Further research is recommended to assess reliability in greater detail as well as the responsiveness of the measure.

**Trial registration** The study is registered in The Netherlands National Trial Register (Trial Code 5860).

## Key Points

The PESaM questionnaire is a unique patient-reported outcome measure developed in conjunction with patients to evaluate experiences and satisfaction with medications.

The generic module of the PESaM questionnaire has sound structural properties and construct validity.

Further research is recommended to assess the reliability and responsiveness of the measure.

The members of the PESaM collaborating group have been listed in the Acknowledgements section.

✉ Merel L. Kimman  
merel.kimman@mumc.nl

Extended author information available on the last page of the article

## 1 Introduction

It is widely acknowledged that the patient's perspective regarding medical therapy should play a central role in deciding upon treatment strategies [1–11]. Patients who perceive their medication to be ineffective, suffer from side effects, or experience difficulties with the administration of a medication, are less likely to be satisfied and take their medication as prescribed [12]. This in turn can impact the effectiveness of the treatment [13], and may result in inefficient use of healthcare resources [14]. Hence, systematic collection of patient experiences is useful in clinical practice, but may also play an important role in health technology assessment (HTA) [1, 9, 15–18]. To date, the patient's views and experiences are increasingly taken into account in the reimbursement decision-making process, mostly through active patient engagement; for example, through consultation rounds with patients or patient representatives [15, 19, 20]. Quantitative data (e.g. questionnaires) of patient experiences, however, have great potential to strengthen the patient evidence that is taken into account [15, 19].

The Patient Experiences and Satisfaction with Medications (PESaM) questionnaire is a recently developed patient-reported outcome measure (PROM) for quantitative assessment of patient experiences and satisfaction with (novel) drug therapies [21]. It was developed in response to the request of two patient organizations that wanted to better capture patients' experience with expensive orphan drugs that had been granted conditional approval for reimbursement by the Dutch government. Conditional approval requires a national registry with data on physiological outcomes as well as evaluation of patient-relevant outcomes such as health-related quality of life (HRQoL) and patient experiences, as input for a re-evaluation after 4 years [22]. The orphan drugs were pirfenidone and nintedanib for the treatment of idiopathic pulmonary fibrosis (IPF) and eculizumab for atypical haemolytic uraemic syndrome (aHUS).

IPF is a chronic and progressive lung disease that is characterized by irreversible loss of lung function. For patients with IPF, prognosis is poor with an average survival of 3–5 years after diagnosis, and treatment options are limited [23]. Pirfenidone and nintedanib are both orally administered anti-fibrotic agents for treatment of mild to moderate IPF. Pirfenidone treatment entails taking three capsules (swallowed whole with water) three times daily with food [24]. Nintedanib treatment consists of taking one capsule twice daily with food, and with 12 h in between administrations [25].

Atypical HUS is an extremely rare and life-threatening disease characterized by suddenly abnormal breakdown

of red blood cells, low platelet counts and acute renal failure due to the formation of blood clots in small vessels, particularly in the kidneys. In the Netherlands, it is estimated that 15–20 people are diagnosed with aHUS each year, of whom five are children. The prognosis for people with aHUS is poor, with around 2–10% of people with the disease dying in the initial, acute phase [26, 27]. Eculizumab is the first and only therapy approved for treatment of aHUS and is administered intravenously. Treatment with eculizumab consists of weekly infusions in the initial phase (up to 4 weeks) followed by eculizumab infusions every 14–21 days depending on the body weight of the patient, for potentially their entire life (according to the product summary of the European Medicines Agency). Following new Dutch guidelines, the interval between eculizumab administrations is gradually extended after 3 months of treatment in patients who are stable and in remission [28].

The focus of the PESaM questionnaire is on a patient's subjective experiences related to the impact of the effectiveness and side effects of the medication on health and daily life, as well as the ease of use of the medication. It comprises two disease-specific modules evaluating drug treatment for IPF and aHUS, a generic module applicable to any medication and a module focused on patient expectations of the medication. The format and content of the PESaM questionnaire were based on a formal conceptual framework [29], a literature review and input from patients through focus groups and individual interviews [21]. Development and pre-testing of the questionnaire has been described extensively elsewhere [21]. The aim of the study described in this paper was to assess the psychometric properties of the questionnaire, with a specific focus on the generic module due to its wide applicability. The items of the generic module evaluate the positive (or negative) influence of a medication's efficacy and side effects on physical health, emotional health and (ability to perform) social activities, and the influence of the medication's mode of administration on daily life. In addition, satisfaction with each domain as well as overall satisfaction with the medication is assessed. The main focus of the psychometric evaluation was to test the construct validity of the measure, that is, the extent to which the scores on the generic module relate to other measures (or constructs) or vary between groups consistent with theoretically derived hypotheses [30]. Based on Strasser and colleagues' model of satisfaction [29], we first tested the association between experiences and satisfaction. Next, due to the measure's emphasis on the impact of the medication on HRQoL indicators, we tested associations between patient experiences and HRQoL, as measured by the EQ-5D, which also focuses on physical, emotional and social health [31]. Finally, it is important that the generic module is able to discriminate between patient groups that are expected

to differ in their experiences and satisfaction, for example because of different therapeutic effects of medications or time on medication. Results of the construct validity and reliability testing are presented and discussed in this paper.

## 2 Study Participants and Methods

### 2.1 Study Participants and Data Collection

The validity study was performed at 11 site hospitals across the Netherlands over a period of 1 year (July 2016–June 2017) using a convenience sample of IPF patients using pirfenidone or nintedanib and aHUS patients receiving eculizumab treatment. In addition, a sample of patients was recruited that had not been involved in the development process of the PESaM questionnaire. Patients using a once-a-day oral formulation of tacrolimus (brand name Advagraf) after kidney transplantation were asked to complete the generic module only.

Inclusion criteria for the validity study were (1) being an adult (aged > 15 years) diagnosed with IPF, aHUS (first diagnosis or recurrence), or a history of kidney transplantation; (2) taking any of the following medications: pirfenidone or nintedanib, eculizumab, or tacrolimus, respectively; and (3)

being able to read Dutch. There were no restrictions to the time on medication or treatment regimen. IPF patients were recruited in nine hospitals across the Netherlands. Atypical HUS patients received their treatment in the Radboud University Medical Center in Nijmegen, the national expertise centre for aHUS. Patients using tacrolimus were attending follow-up visits in the Maastricht University Medical Centre. The appropriate modules of the PESaM questionnaire were either handed out (on paper) to patients during outpatient clinic follow-up visits or emailed via an online survey system.

Written informed consent was obtained prior to completing the questionnaire. The study protocol for the development and testing of the measure was reviewed and approved by the Medical Ethics Committee of Erasmus Medical Centre (MEC-2015-265). The study is registered in The Netherlands National Trial Register (trial code 5860).

### 2.2 Measures

#### 2.2.1 Generic Module of the PESaM Questionnaire

The generic module of the PESaM questionnaire consists of 18 items related to the domains of effectiveness (4 items), side effects (4 items), ease of use (3 items) and satisfaction

**Table 1** Domains, items and response options of the generic module of the PESaM questionnaire

Domain	Items	Response options
Effectiveness	Efficacy of medication	No positive effectiveness to very positive effectiveness (5 levels) I don't know
	Positive influence on physical health (e.g. walking, cycling)	No positive influence to very positive influence (5 levels)
	Positive influence on feelings and emotions (e.g. fear, joy)	I don't know
	Positive influence on social activities (e.g. work, family, friends)	
	Satisfaction (or dissatisfaction) with effectiveness	Thermometer – 5 (very dissatisfied) to + 5 (very satisfied)
Side effects	Bothersomeness of side effects	I did not experience side effects (continue to item 10) Not bothersome to very bothersome (5 levels) I don't know
	Negative influence on physical health (e.g. walking, cycling)	No negative influence to very negative influence (5 levels)
	Negative influence on feelings and emotions (e.g. fear, joy)	I don't know
	Negative influence on social activities (e.g. work, family, friends)	
	Satisfaction (or dissatisfaction) with side effects	Thermometer – 5 (very dissatisfied) to + 5 (very satisfied)
Ease of use	Inconvenience of administration mode	Not inconvenient to very inconvenient (5 levels)
	Inconvenience of time table (frequency)	
	Inconvenient to incorporate in everyday life	
	Satisfaction (or dissatisfaction) with ease of use	Thermometer – 5 (very dissatisfied) to + 5 (very satisfied)
Overall satisfaction	Overall (dis) satisfaction with medication	Thermometer – 5 (very dissatisfied) to + 5 (very satisfied)
	Importance of effectiveness	Not important to very important (5 levels)
	Importance of effectiveness	
	Importance of effectiveness	

(4 items) (Table 1). The final three items concern the perceived importance of the effectiveness, side effects and ease of use of the medication. These items were included for validation purposes only. Items within the domains of effectiveness and side effects relate to the impact of the medication on aspects of physical health, emotional health and daily life (social and work activities) [21]. These items all relate to subjective experiences of the respondent (e.g. to what extent did the side effects of the medication negatively influence your daily life?). The ease-of-use items focus on the administration mode and the (potential) interference of administering the medications with daily life. Each domain has one item on satisfaction (e.g. how satisfied or dissatisfied were you with the ease of use of the medication?). In addition, there is an ‘overall satisfaction’ item asking the respondent to take all domains of the medication into account. A 5-point (Likert-type) scale with the following anchor levels was chosen as the response format for items evaluating experiences: ‘not at all’, ‘a little’, ‘reasonable’, ‘a lot’ and ‘very’. In the domains of effectiveness and side effects, a ‘don’t know’ response category was added. Items regarding satisfaction were scored using a (horizontal) thermometer, ranging from –5 (very dissatisfied) to +5 (very satisfied).

Mean scores for experiences in each domain of the generic module were calculated if at least two items of the domain were completed. Response categories were coded from 0 (‘not at all’) to 4 (‘very’) and the domain scores can therefore range between 0 and 4. The ‘don’t know’ response category was considered a missing value for the purpose of calculating mean scores. Higher scores in the domain of effectiveness represented higher positive experiences regarding the effectiveness of the medication, while in the domains of side effects and ease of use, higher scores represented higher burden from side effects and lower ease of use, respectively. To facilitate interpretation, original scores in the domains of side effects and ease of use were recoded so that higher scores represented low burden of side effects and ease of use (i.e. for all domains, higher scores represent positive experiences). The items related to satisfaction (i.e. items 5, 10, 14 and 15) were not recoded; the reported scores (ranging between –5 and 5) were used.

### 2.2.2 Disease-Specific Modules of the PESaM Questionnaire

IPF and aHUS patients completed the applicable disease-specific module of the PESaM questionnaire in conjunction with the generic module. Similar to the generic module, the disease-specific modules focus on experienced effectiveness, side effects and ease of use of the medication, but do not include items regarding satisfaction. The items on effectiveness assume a positive influence on health, the items on side effects a negative influence and the items on ease of use

focus on the potential inconvenience of the specific mode of administration and whether patients have skipped medication. The difference between these and the generic module is that the disease-specific modules evaluate the effectiveness of the medication on specific disease symptoms. For example, the module for aHUS asks about the influence of eculizumab on energy levels, the ability to participate in society and fear of infection (meningitis), and provides a checklist of potentially experienced side effects. The module for IPF focuses on its perceived ability to slow down disease progression, reduce coughing, feeling tired, out of breath and whether respondents experienced side effects such as photosensitivity, nausea and diarrhoea. Details on the contents and response levels (e.g. which side effects are included) can be found in our earlier paper [21]. Response levels and the scoring of domain scores for effectiveness and ease of use are similar to the generic module and range between 0 and 4. For each experienced side effect, respondents are asked to rate how bothered they are by that side effect and a sum score for this domain is calculated by multiplying the number of side effects with the respondent’s average rating of bothersomeness. Thus, a higher score represents a higher level of bothersomeness experienced due to one or more side effects.

### 2.2.3 EQ-5D

Participants completed the EQ-5D (3-level version) to measure HRQoL [31]. The EQ-5D consists of two components: a descriptive system of health and a visual analogue rating scale (VAS). The descriptive system consists of five items (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), each with three response levels (no problems, some problems and severe problems). Reported health states on the descriptive system are converted to an EQ-5D index score where 1 represents full health and 0 represents death [32]. The EQ-VAS records the respondent’s self-rated health on a 0–100 scale, with the endpoints respectively labelled ‘worst imaginable health state’ and ‘best imaginable health state’.

### 2.2.4 Demographic and Clinical Data

Date of birth, gender, date of diagnosis, medication and time on medication were collected from electronic medical files. For IPF patients, forced vital capacity (FVC) at diagnosis and around completion date of the questionnaire were collected, expressed in percentage and litres.

## 2.3 Data Analysis

Due to a limited number of repeated measurements, a cross-sectional dataset was used for testing psychometric

properties of the generic module. In case participants had completed the questionnaires at multiple time-points during the study, only their first completed questionnaire was used for this analysis.

### 2.3.1 Structural Properties

A stacked bar chart graphically presents the distribution of scores, percentage of missing values, and ‘don’t know’ responses. Floor and ceiling effects were considered present when at least 15% of respondents scored the lowest or highest possible score, respectively [33].

### 2.3.2 Internal Consistency

Internal consistency as a measure of the extent to which the items in the domains of effectiveness, side effects and ease of use are correlated (homogeneous), thus measuring the same concept, was assessed by calculating the Cronbach’s alpha for each domain separately. A low Cronbach’s alpha indicates a lack of correlation between the items, which makes summarizing the items unjustified. The internal consistency was considered good when Cronbach’s alpha was between 0.70 and 0.95 [30].

### 2.3.3 Construct Validity

Confirmatory factor analysis (CFA) with robust maximum likelihood estimation was used to test the factor structure of the generic module. We hypothesized that items 1–4 measure the latent construct ‘experienced effectiveness’, items 6–9 ‘bothersomeness of side effects’, and items 11–13 represent ‘ease of use’. First, goodness-of-fit indices were used to evaluate the adequacy of the model’s fit to the data, including the Chi-square value, comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA) accompanied by its 90% confidence interval (CI) and standardized root mean square residual (SRMR). CFI and TLI values exceeding 0.95, and SRMR and RMSEA values close to (or less than) 0.08 and 0.06, respectively, represented a good fit [34]. Second, convergent validity was examined by assessing the size and significance of the factor loadings (using the standardized regression coefficients) and the average variance extracted (AVE) for each factor, which should exceed 0.50 [35]. Finally, discriminant validity (i.e. whether each one of the domains has enough discriminant validity from the other domains) was evaluated and considered good when the correlation coefficients between the three domains were  $< 0.8$ .

Construct validity was further assessed by investigating whether scores on the domains (i.e. constructs) of the PESaM generic module (e.g. experiences and satisfaction) relate to other constructs in a manner that is consistent with

a priori hypotheses [30, 36]. The following hypotheses were tested:

1. It was a priori expected that patient experiences have a medium to high correlation with satisfaction; more positive experiences (e.g. low burden of side effects) are associated with higher levels of satisfaction [4, 29, 37].
2. It was expected that the effectiveness of the medication was considered the most important domain of medication use to patients and therefore that ‘satisfaction with effectiveness’ and ‘experiences with effectiveness’, respectively, have the largest independent contribution to overall satisfaction, relative to the contributions of the other domains (i.e. side effects and ease of use).
3. Positive patient experiences in the PESaM refer to an experienced positive impact on physical, emotional and social health, respectively. It was therefore expected that patient experience scores for the domains effectiveness and side effects of the medication were moderately correlated with HRQoL [4, 38, 39].
4. Patient experiences of IPF and aHUS patients regarding effectiveness, side effects, and ease of use, as reported in their respective disease-specific modules, were expected to have a strong (positive) correlation with the corresponding experiences as reported in the generic module.

The strength and direction of the associations were measured using the Pearson product-moment correlation or Spearman’s rank-order correlation coefficient, depending on the distribution of the mean scores or measurement scale (hypotheses 1, 3 and 4). A correlation coefficient between 0.90 and 1.00 would indicate a very high positive correlation, 0.70–0.90 a high (positive) correlation, 0.50–0.70 a moderate (positive) correlation, 0.30–0.50 a low (positive) correlation, and 0–0.30 a negligible correlation [40].

Multiple regression analyses (enter method) were conducted to examine the relationship between overall satisfaction (item 15) and several proposed predictors (hypothesis 2). Unstandardized regression coefficients (B) and explained variance (adjusted  $R^2$ ) were estimated for two multiple regression models: (i) satisfaction with effectiveness (item 5), satisfaction with side effects (item 10), and satisfaction with ease of use (item 4) as predictors of overall satisfaction; and (ii) experiences with effectiveness (domain score), experiences with side effects (domain score), and experiences with ease of use (domain score) as predictors of overall satisfaction. Mean scores on the patients’ rating of importance of each domain (items 16, 17 and 18 of the generic module) were compared using repeated measures ANOVA and post hoc testing (Bonferroni). Subsequently, the ranking of the importance of the domains (in case of significant differences between mean scores) were compared with the results of the regression



models (i.e. the size and significance of the independent contribution of a domain to overall satisfaction). It was hypothesized that a direct measurement of the importance of domains (items 16–18) would produce a ranking that is similar to results of the regression analyses (where the domain with the largest regression coefficient is considered to be the most important domain to the patient).

Known-groups validity (i.e. whether an instrument shows different scores for groups that in theory should have different scores) was tested by comparing mean scores of the different medications.

We expected varying scores between the drugs for different diseases due to varying therapeutic effects, side effects and modes of administration. More specifically, we tested the following hypotheses:

1. It was expected that patient experiences regarding effectiveness were more positive for eculizumab and tacrolimus compared with pirfenidone or nintedanib, since the latter therapies aim to reduce disease progression, which may be difficult to experience by patients over short periods (i.e. ‘the past 4 weeks’).
2. It was expected that eculizumab users report less positive experiences with ease of use, due to the intravenous administration requiring inpatient hospital visits.
3. No major differences in ease of use were expected between pirfenidone, nintedanib and tacrolimus, all requiring oral administration.
4. A final hypothesis is that long-term users of a medication (i.e. > 2 months on medication) had more positive experiences and were more satisfied with a medication compared with new users (i.e. ≤ 2 months) [37].

Analysis of variance (ANOVA) with post hoc (Tukey) testing was used for comparing scores between medications. The independent sample *t* test was used when two groups (i.e. new users vs long-term users) were compared. A *p* value of 0.05 was used as the cut-off for significance.

### 2.3.4 Reliability

A subsample of patients in stable health, as determined by their healthcare provider, were asked to complete the generic module a second time, 2 weeks after their initial completion (test–retest). The Intraclass Correlation Coefficient (ICC) was used to test the reliability of the questionnaires between the two measurements [41]. ICC estimates and their 95% confident intervals were calculated based on absolute agreement and a 2-way mixed-effects model. Following recommendations by Terwee et al., the reliability is positively rated when the ICC is at least 0.70 [30].

Statistical analyses were performed in SPSS statistical package version 23 (IBM SPSS, IBM) and R version 3.5.1 using the lavaan package [42].

## 3 Results

A total of 188 patients (48% pirfenidone, 36% nintedanib, 11% tacrolimus, 5% eculizumab) completed the generic module of the PESaM questionnaire. Of these patients, 116 also completed (all items of) the EQ-5D, 159 completed the disease-specific module (93% IPF and 7% aHUS), and 39 completed the generic module twice for test–retest measurement purposes. The median time on medication was 9 months (range 1–230 months). Characteristics of the study sample are presented in Table 2.

### 3.1 Structural Properties

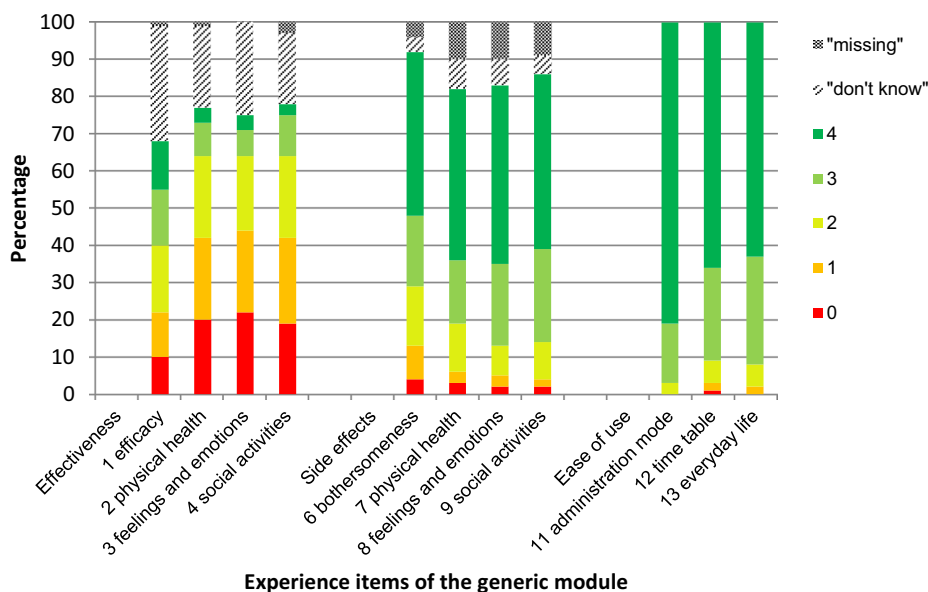
The distribution of response levels was satisfactory for most items of the generic module (see Fig. 1). All items, except items 11 and 13, showed responses across the full range of response options. There were < 1% missing responses on the items in the domains of effectiveness and ease of use. Within the domain of side effects, there were between 3 and 10% missing responses on individual items. Highest use (31%) of the ‘don’t know’ category was for item 1 of the questionnaire: “to what extent did you experience a positive effect of the medication?” Responses on the items related

**Table 2** Patient characteristics of the validity study

Patient characteristics	Pirfenidone ( <i>n</i> = 91)	Nintedanib ( <i>n</i> = 67)	Eculizumab ( <i>n</i> = 10)	Tacrolimus ( <i>n</i> = 20)	All ( <i>n</i> = 188)
Mean age in years (min–max)	72 (52–86)	72 (58–85)	42 (26–72)	63 (37–76)	69 (26–86)
Sex, male (%)	75 (82)	54 (81)	4 (40)	9 (45)	142 (76)
Time since diagnosis in months, median (min–max)	15 (1–122)	12 (1–78)	74 (1–153)	87 (12–230)*	16 (1–230)
Time on medication in months, median (min–max)	12 (1–57)	5 (0–48)	7 (0–15)	87 (12–230)	9 (0–230)

\*Time since kidney transplant

**Fig. 1** Stacked bar chart of response distribution of the experience items in the generic module of the PESaM questionnaire ( $n = 188$ ). Responses presented are the recoded scores so that high scores indicate positive experiences (e.g. a score of 4 represents 'very positive influence', 'no negative influence' and 'very convenient' for the domains effectiveness, side effects and ease of use, respectively)



to satisfaction generally covered the full range of response categories (– 5 to 5) and there were few missing responses (< 1%). The mean scores were 2.2 (SD 2.4), 1.9 (SD 2.8), 3.3 (SD 1.9) and 2.6 (SD 2.3) for the satisfaction items in the domains of effectiveness (item 5), side effects (item 10), ease of use (item 14), and overall satisfaction (item 15), respectively.

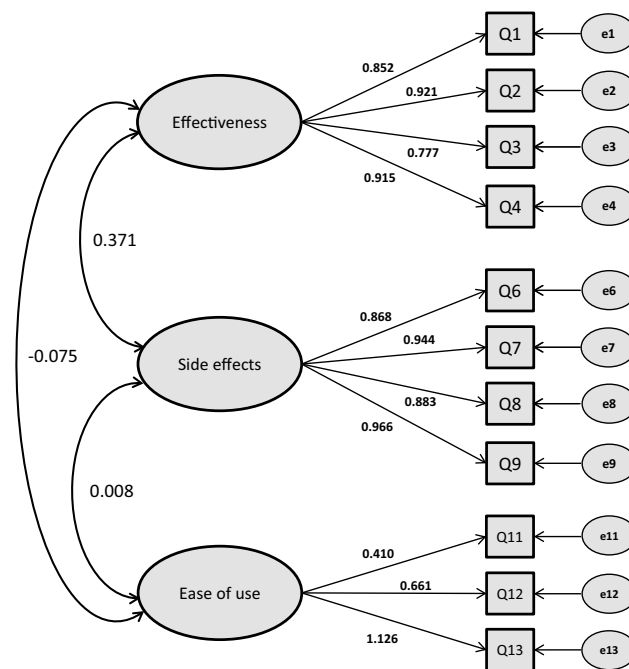
Floor and ceiling effects were considered present in an item when at least 15% of respondents scored the lowest or highest possible score, respectively. A floor effect was identified in items 2, 3 and 4 of the effectiveness domain; for each item, approximately 20% of respondents did not experience any positive effect of the medication. In the side effects domain, between 40 and 50% of the respondents reported the highest possible score on each item, and for items in the domain for ease of use, the ceiling effect was significant with approximately 60% and 81% of respondents reporting the highest level (i.e. no problems with ease of use).

**3.2 Internal Consistency**

The internal reliability of the domains of effectiveness, side effects, and ease of use was examined. The Cronbach’s alpha values were 0.92, 0.93 and 0.80, respectively, and fell within the recommended range of 0.70–0.95, providing evidence for internally consistent (homogeneous) scales [30].

**3.3 Construct Validity**

Figure 2 is a diagrammatic representation of the CFA of the generic module of the PESaM questionnaire. Listwise deletion of cases missing any variable in the dataset resulted in



**Fig. 2** Results of the confirmatory factor analysis

86 used observations (out of a total of 188). The Chi-square statistic was not significant (Chi-square value 37.94,  $df = 41$ ,  $p = 0.607$ ), representing an acceptable model. The RMSEA was zero (90% CI 0.000–0.065) and both the CFI value of 1.000 and TFI value of 1.006 exceeded 0.90, suggesting a model with satisfactory fit. Convergent validity was considered good, with moderate to large sizes of all factor loadings (all significant with  $p < 0.001$ ) (Fig. 2). In addition, the AVE

for effectiveness, side effects and ease of use were 0.753, 0.595 and 0.624, respectively, and were considered satisfactory (i.e. values  $> 0.50$ ). Finally, correlation coefficients between the three hypothesized factors (i.e. effectiveness, side effects and ease of use) ranged from  $-0.075$  ( $p = 0.454$ ) to  $0.371$  ( $p = 0.003$ ) and were well below the 0.80 threshold value supporting discriminant validity between the domains.

In line with the first hypothesis, there was a significant moderately positive association between patient experiences and satisfaction for the domains of effectiveness and side effects (Spearman's Rho 0.699 and 0.625, respectively). However, there was a low association between experiences and satisfaction for the ease of use domain (Spearman's Rho 0.475). Second, we expected there to be a low to moderate correlation between experiences with effectiveness and side effects (thus their impact on health) and HRQoL. Contrary to this expectation, positive experiences with effectiveness were not associated with a better HRQoL as represented by a higher EQ-5D index score (Pearson's  $r = 0.175$ ,  $p = 0.100$ ). Patient experiences with effectiveness did have a low correlation with HRQoL as reported on the VAS (Pearson's  $r = 0.397$ ,  $p < 0.001$ ). Positive experiences with side effects (i.e. a low burden and impact on health) were found to have a low, but significant, correlation with the EQ-5D index and the VAS (Pearson's  $r = 0.386$ ,  $p < 0.001$  and Pearson's  $r = 0.271$ ,  $p = 0.007$ , respectively). Finally, strong associations were found between experiences with effectiveness, side effects and ease of use as reported on the generic module of the PESaM questionnaire and corresponding domains on the disease-specific modules (data not shown). For example, the score of IPF patients on the domain experiences with effectiveness of the generic module, represented by general items such as experienced impact on physical health and impact on daily life, was highly correlated with the score for patient 'experiences with effectiveness' as measured by the disease-specific module that includes items on disease progression, cough and fatigue to evaluate effectiveness (Spearman's Rho 0.872,  $p < 0.001$ ). Correlation coefficients for side effects and ease of use between the generic and disease-specific modules were 0.617 and 0.782 (all  $p < 0.001$ ), respectively.

Regression analyses identified the item 'satisfaction with effectiveness' as the strongest predictor of 'overall satisfaction' ( $B = 0.63$ ,  $p < 0.001$ ), relative to the contribution of 'satisfaction with side effects' ( $B = 0.20$ ,  $p < 0.001$ ) and 'satisfaction with ease of use' ( $B = 0.16$ ,  $p = 0.049$ ), with the model explaining 70% of the variance (adjusted  $R^2 = 0.70$ ). In the second regression model that examined the impact of the domains of patient experiences on overall satisfaction, 'experiences with effectiveness' was identified as the strongest predictor of overall satisfaction ( $B = 1.25$ ,  $p < 0.001$ ), relative to the contribution of 'experiences with side effects' ( $B = 0.81$ ,  $p < 0.001$ ) and 'experiences with ease

of use' ( $B = 0.29$ ,  $p = 0.263$ ). The performance of this model was reasonable, explaining 50% of the variance (adjusted  $R^2 = 0.50$ ).

Results of the regression models were reasonably in line with the scores reported on items 16, 17 and 18 of the generic module where the effectiveness of medication was found to be of highest importance to respondents (mean score of 3.6 with SD 0.8), followed by ease of use with a mean score of 2.5 (SD 1.3) and side effects with a mean score of 2.4 (SD 1.3). Repeated measures ANOVA showed that the difference between the mean importance scores of the domains were statistically significant ( $F(2, 362) = 92.498$ ,  $p < 0.001$ ). Post-hoc tests using the Bonferroni correction revealed that there was a significant difference between item 16 (effectiveness) and item 17 (side effects) ( $p < 0.001$ ) and item 16 and item 18 (ease of use) ( $p < 0.001$ ). However, the mean importance scores for items 17 and 18 were not significantly different ( $p = 0.268$ ). Thus, while ease of use received a score as high as side effects when patients were asked to rate the importance of the medication characteristic directly, regression models showed that, relative to the domains of effectiveness and side effects, ease of use experiences contributed the least to overall satisfaction with the medication.

### 3.3.1 Known-Groups Validity

We observed varying scores for different patient groups that completed the PESaM questionnaire. As hypothesized, patients using eculizumab or tacrolimus reported more positive experiences regarding effectiveness of the medication than patients using pirfenidone or nintedanib (Table 3). The higher score for tacrolimus (mean 2.5) compared with pirfenidone (mean 1.1) and nintedanib (mean 1.5) was statistically significant ( $p = 0.001$  and  $p = 0.008$ , respectively). However, the score for eculizumab (mean 2.2) was not significantly different to the mean scores for pirfenidone and nintedanib, with  $p = 0.132$  and  $p = 0.302$ , respectively. This may be explained by the small sample of eculizumab users.

Secondly, as hypothesized, patients using eculizumab generally reported less positive experiences and lower satisfaction scores in the domain 'ease of use' compared with the oral therapies. Nevertheless, except for the difference in satisfaction between eculizumab (mean 2.0) and tacrolimus (mean 4.3) ( $p = 0.022$ ), the differences between the medications lack statistical significance ( $p = 0.083$  and  $p = 0.080$  for the differences with pirfenidone and nintedanib, respectively). Third, as expected, patient experiences in the domain 'ease of use' for the three oral medications were indeed very similar, with the exception that the satisfaction score for ease of use was significantly higher for tacrolimus (mean 4.3, SD 1.1) compared with pirfenidone and nintedanib (mean 3.3, SD 1.7,  $p = 0.016$  and mean 3.3, SD 2.0,  $p = 0.022$ , respectively).



**Table 3** Mean scores (SD) for pirfenidone, nintedanib, tacrolimus and eculizumab

Domains	Pirfenidone <sup>a</sup> , <i>n</i> = 76	Nintedanib <sup>b</sup> , <i>n</i> = 49	Tacrolimus <sup>c</sup> , <i>n</i> = 14	Eculizumab <sup>d</sup> , <i>n</i> = 7	ANOVA	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	<i>F</i>	<i>p</i> value
Experiences with effectiveness <sup>e</sup>	1.3 (1.1)	1.5 (1.0)	2.5 (0.9)	2.2 (1.1)	<i>F</i> (3.150) = 6.3	< <b>0.001</b>
Experiences with side effects <sup>e</sup>	3.0 (1.0)	3.3 (0.9)	3.8 (0.4)	2.4 (1.4)	<i>F</i> (3.157) = 6.1	<b>0.001</b>
Experiences with ease of use <sup>e</sup>	3.6 (0.5)	3.6 (0.6)	3.8 (0.4)	3.2 (0.7)	<i>F</i> (3.183) = 2.8	<b>0.040</b>
Satisfaction with effectiveness <sup>f</sup>	1.9 (2.5)	1.9 (2.2)	3.7 (1.7)	3.7 (1.6)	<i>F</i> (3.182) = 5.4	<b>0.001</b>
Satisfaction with side effects <sup>f</sup>	1.6 (2.9)	2.0 (2.8)	3.2 (1.9)	1.8 (3.0)	<i>F</i> (3.180) = 1.7	0.165
Satisfaction with ease of use <sup>f</sup>	3.3 (1.7)	3.3 (2.1)	4.3 (1.1)	2.0 (2.9)	<i>F</i> (3.184) = 3.4	<b>0.019</b>
Overall satisfaction <sup>f</sup>	2.4 (2.4)	2.4 (2.3)	4.1 (1.5)	3.5 (1.9)	<i>F</i> (3.182) = 3.8	<b>0.012</b>

The significant *p* values are in bold (*p* ≤ 0.05)

<sup>a</sup>The number of respondents vary between 76 and 91 because a domain score could not always be calculated (e.g. when > 2 responses were missing or 'don't know')

<sup>b</sup>The number of respondents vary between 49 and 58

<sup>c</sup>The number of respondents vary between 14 and 20

<sup>d</sup>The number of respondents vary between 7 and 10

<sup>e</sup>Domain scores range between 0 and 4, with higher scores representing more positive experiences

<sup>f</sup>Satisfaction scores range between - 5 (very dissatisfied) and 5 (very satisfied)

**Table 4** Mean scores (SD) for new (< 2 months) and long-term (≥ 2 months) users

Domains	New users <sup>a</sup> , <i>N</i> = 26–33	Long-term users <sup>a</sup> , <i>N</i> = 128–153	Independent <i>t</i> test	
	Mean (SD)	Mean (SD)	<i>t</i>	<i>p</i> value
Experiences with effectiveness <sup>b</sup>	1.0 (0.8)	1.6 (1.1)	<i>t</i> (152) = - 2.7	<b>0.007</b>
Experiences with side effects <sup>b</sup>	3.2 (1.2)	3.2 (0.9)	<i>t</i> (159) = 0.1	0.942
Experiences with ease of use <sup>b</sup>	3.6 (0.5)	3.6 (0.5)	<i>t</i> (185) = - 0.5	0.630
Satisfaction with effectiveness <sup>c</sup>	1.0 (2.5)	2.4 (2.3)	<i>t</i> (184) = - 3.1	<b>0.003</b>
Satisfaction with side effects <sup>c</sup>	1.3 (2.9)	2.0 (2.8)	<i>t</i> (182) = - 1.2	0.217
Satisfaction with ease of use <sup>c</sup>	2.8 (2.2)	3.4 (1.8)	<i>t</i> (186) = - 1.9	0.061
Overall satisfaction <sup>c</sup>	1.9 (2.6)	2.8 (2.2)	<i>t</i> (184) = - 2.1	<b>0.042</b>

The significant *p* values are in bold (*p* ≤ 0.05)

<sup>a</sup>The number of respondents vary because a domain score could not always be calculated (e.g. when > 2 responses were missing or 'don't know')

<sup>b</sup>Domain scores range between 0 and 4, with higher scores representing more positive experiences

<sup>c</sup>Satisfaction scores range between - 5 (very dissatisfied) and 5 (very satisfied)

Finally, higher mean satisfaction scores for long-term users compared with new users were observed across all domains, although the difference in satisfaction of side effects did not reach statistical significance (Table 4). Interestingly, although satisfaction levels were generally higher, the experienced ease of use of the medication did not differ between new and long-term users.

The patient groups differed regarding type of medication (and thus administration mode, therapeutic effect and side effects), but also in time on medication, severity of disease and probably other patient characteristics that were not measured in this study. Score differences should therefore

only be interpreted for validity testing and not be used for a direct comparison between medications.

### 3.4 Reliability

A test–retest measurement was conducted in a subgroup of 39 patients, of whom 18 used pirfenidone, six used nintedanib and 15 used tacrolimus. The mean age of the patients was 68 years old (range 37–86) and 69% were male. The average time on treatment was 35 months (range 1 month–27 years). The mean time on treatment for tacrolimus patients was 6 years (range 1–24 years), while for IPF it was 12 months (range 1–57 months). The ICCs for single

**Table 5** Intraclass correlations (ICC) for each item at T1 and T2 ( $n = 39$ )

Item	T1, mean (SD)	T2, mean (SD)	ICC <sup>a</sup> , (95% CI)
<i>Effectiveness</i>			
Efficacy of medication	2.7 (1.3)	2.5 (1.3)	0.57 (0.22–0.79)
Physical health	1.9 (1.3)	2.0 (1.1)	0.71 (0.47–0.86)
Feelings and emotions	1.7 (1.2)	1.6 (1.2)	0.59 (0.31–0.78)
Social activities	1.9 (1.2)	2.1 (1.1)	0.71 (0.47–0.86)
<i>Domain score effectiveness</i>	2.0 (1.7)	1.9 (1.1)	0.76 (0.57–0.88)
Satisfaction (or dissatisfaction)	2.5 (2.4)	2.7 (2.0)	0.66 (0.43–0.81)
<i>Side effects</i>			
Bothersomeness side effects	3.5 (0.8)	3.2 (1.1)	0.61 (0.34–0.78)
Physical health	3.6 (0.8)	3.6 (0.7)	0.52 (0.21–0.84)
Feelings and emotions	3.8 (0.6)	3.6 (0.7)	0.71 (0.49–0.84)
Social activities	3.7 (0.7)	3.7 (0.6)	0.50 (0.18–0.84)
<i>Domain score side effects</i>	3.6 (0.7)	3.5 (0.7)	0.75 (0.56–0.87)
Satisfaction (or dissatisfaction)	3.6 (0.7)	3.5 (0.7)	0.72 (0.53–0.84)
<i>Ease of use</i>			
Administration mode	3.9 (0.3)	3.9 (0.4)	0.39 (0.09–0.62)
Time table (frequency)	3.7 (0.7)	3.7 (0.5)	0.42 (0.12–0.65)
Incorporate in everyday life	3.7 (0.7)	3.7 (0.5)	0.18 (-0.15–0.47)
<i>Domain score ease of use</i>	3.8 (0.4)	3.8 (0.4)	0.48 (0.19–0.69)
Satisfaction (or dissatisfaction)	4.0 (1.3)	3.8 (1.7)	0.47 (0.18–0.68)
<i>Overall satisfaction</i>			
Overall (dis) satisfaction	3.3 (2.0)	3.3 (1.8)	0.76 (0.59–0.87)

<sup>a</sup>ICC for absolute agreement

items were generally classified as moderate and the ICCs for domain scores as fair. Items regarding ease of use and the subsequent domain score show very low ICCs suggesting poor reliability (Table 5). Visual inspection of the data showed that individual scores were almost identical for the two measurements, but little variation and a couple of outliers were likely to cause low and unreliable values for the domain ‘ease of use’ [41].

## 4 Discussion

It is increasingly acknowledged that the patient’s perspective is of great importance when assessing the value of healthcare technologies. The generic module of the PESaM questionnaire was specifically developed to evaluate the perceived effectiveness, side effects and ease of use of medications, and how these experiences impact on a patient’s health and daily life. Here we report for the first time on the psychometric properties of the PESaM questionnaire. This study showed that the generic module of the PESaM questionnaire is easy to complete and has good construct validity. The reliability of the measure is rated as moderate to fair and requires further investigation.

Structural properties of the generic module were satisfactory, with good response distribution and few missing

items. Of note was the relatively high proportion (19–31%) of respondents that used the ‘don’t know’ response category in the effectiveness domain. The items may have been too general or difficult for respondents who have a disease such as IPF (84% of the sample). Inherent to the nature of the disease, the aim of treatment is to slow down lung function decline, which is often not subjectively experienced by the patient. Moreover, interviews held with IPF patients revealed that patients may find it difficult to discern the effects of pirfenidone or nintedanib from the effects of other medications and treatments they were receiving (e.g. physical rehabilitation or medications to treat side effects and co-morbidities), possibly leading them to a ‘don’t know’ response [21]. Since the PESaM questionnaire explicitly links the evaluated patient outcomes (experiences and satisfaction) to the intervention (medication), the ‘don’t know’ response category was added to ensure that respondents do not feel forced into a response category they cannot relate to. Another notable finding was the floor and ceiling effects identified in the study. Floor and ceiling effects may be considered a threat to validity due to the loss of ability to detect improvements or deteriorations. Ceiling effects were most prominent in the domains of ease of use and side effects. However, in this case, ceiling effects may be due to the specific sample used. The majority of the study sample used easy-to-administer tablets, two to three times a day, and had been using the

medication for some time. It is likely that the sample had become used to incorporating the medication into daily life, reporting little to no problems with ease of use. Furthermore, a large proportion of the respondents were not experiencing side effects at the time of measurement (on average 9 months on medication), which resulted in the maximum (positive) score for this domain. The ceiling effect present in the domain 'ease of use' may have also impacted the test–retest measurement, since the magnitude of an ICC decreases the more similar participants score to each other as a group [43]. Future studies, including medications administered via other routes, should confirm this rationale.

Confirmatory factor analysis provided evidence of the construct validity of the measure. Good model fit suggests that the data adequately represent the underlying theory of the measure. Furthermore, low correlations between the domains and high factor loadings support discriminant and convergent validity of the measure. Since the internal consistency of the domains was also rated as very good, the domains appear to be sufficiently homogeneous for their items to be pooled.

Most hypotheses regarding associations with similar or dissimilar constructs were met. Following Strasser's conceptual model of patient satisfaction, experiences and satisfaction are related but quite distinct concepts, with multiple factors influencing the strength of their association, such as a patient's expectations, beliefs and value system [29]. For example, a respondent may report a negative impact of side effects on daily life, but may not be dissatisfied because these side effects were expected. This study confirmed a positive association between experiences and satisfaction that was only of moderate strength. Furthermore, findings showed that while experiences were sometimes not very different between groups, for example between new and long-term users of medications, satisfaction scores were significantly higher among long-term users. A major strength of the PESaM questionnaire is thus that it specifically distinguishes between the patient's experiences and the patient's satisfaction. Other generic measures of patient experiences with medication, such as the Treatment Satisfaction Questionnaire for Medication (TSQM) [37] and Treatment Satisfaction with Medicines Questionnaire (SATMED-Q) [44], generally have a stronger focus on satisfaction, with no means to separate the two concepts.

Contrary to expectations, positive experiences with effectiveness were not associated with a better HRQoL as measured by the EQ-5D, and the association between patient experiences with side effects and HRQoL was rather weak. It may be that our hypothesis was incorrect and that actually patients with worse HRQoL have more opportunity to benefit from the medications and thus more positive experiences resulting in a negative relation (or at least a blurred relationship due to patients with different health

states). Longitudinal data are better suited to explore the relationship between HRQoL and experiences. This finding does emphasize that HRQoL measures alone, as a way to incorporate the patient's perspective, are insufficient to fully grasp the patient's experience of a drug therapy. We observed the correlation between the generic and disease-specific measure to be very strong, however, which encourages the use of the shorter generic module in both research and clinical settings. The generic module can promote more coherent outcome reporting across studies investigating the effectiveness of medications that aim to include an evaluation of patient experiences.

Finally, the PESaM scores generally adhered to expected patterns across the therapeutic effectiveness of medications and their mode of administration and were able to distinguish between patients who just started on their medication and long-term users. These findings further support the overall validity of the tool.

A number of other limitations related to the study design need to be considered. First, the generic module was tested in the same disease population that was involved in the development of the measure. While a small sample of patients with a different disease were involved (i.e. patients using tacrolimus), the majority of the study population were IPF and aHUS patients. Extrapolation of the results to other medications or diseases must therefore be performed with due caution. Furthermore, the lack in variety of medications impacted on the validity tests. It was difficult to assess, for example, whether ceiling effects identified in the study were due to the metric properties of the questionnaire or due to the nature of the disease and medication of the majority of the study population (i.e. IPF). While we attempted to include participants using other medications, only small samples for eculizumab and tacrolimus were obtained. Due to the extreme rarity of aHUS and hence the use of eculizumab, and the small sample of tacrolimus patients, separate validity tests (e.g. multiple comparison testing as part of ANOVA) for these groups were hindered by a lack of power. In this study, however, the development of the PESaM questionnaire was a specific response to patient organizations and clinicians interested in collecting patient evidence of novel drug therapies in this population of patients with rare diseases. Reliability testing was also affected by the relatively high proportion of 'don't know' responses, which were to be considered as a missing value for calculation of ICCs (which is not suitable for categorical variables). This reduced the sample size for items with this response category available (i.e. items in the domains of effectiveness and side effects) resulting in a sample that may be considered too small for proper reliability analysis. Further investigation of the reliability of the measure is recommended. Unfortunately, due to logistic reasons, insufficient data were available to assess the responsiveness of the measure. While a proportion of

IPF patients completed the measure at multiple time points, FVC and HRQoL did not change sufficiently between measurements (typically 3 months apart) to identify groups of patients with varying degrees of change. A final limitation is that this primary validation study enrolled patient samples obtained in the Netherlands; consequently, validation of this instrument in international settings is unknown and must be tested.

Further use of the PESaM questionnaire in clinical studies, accompanied by validity research, is recommended to assess the validity in a variety of patient groups with drug therapies that differ in terms of administration modes and therapeutic effects, as well as the reliability of the measure. Longitudinal data are needed to support testing of the responsiveness and to identify the minimally important difference of the measure. Furthermore, the relationship between patient experiences and issues such as medication adherence and switching (e.g. behavioural response), as well as the potential of the PESaM questionnaire to assist shared decision making in clinical practice, are areas of interest for future research.

## 5 Conclusions

The PESaM questionnaire is a unique patient-reported outcome measure evaluating patient experiences and satisfaction with medications. It has been developed in response to a request from patient organizations and clinicians, and in conjunction with patients, ensuring coverage of domains and issues relevant from the patient's perspective. Results of this first validity study are promising, with the generic module showing sound structural properties, internal consistency, and construct validity. Further research is encouraged to examine the reliability and responsiveness in more detail and assess the generalizability of the results to support broader implementation of the measure.

**Data Availability Statement** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

**Acknowledgements** MK, MW, KW, NvdK, MS, XvJ and CD contributed to the study design. PB, IC, ND, BvD, JG, MG, KdH, RJ, BK, HK, MvM, JM, KM, RM, EN, MO, EP, HvR, HR, NT, AvV, LV, GdV, MW, IW and JW recruited participants and collected data. SvK contributed to the statistical analyses. MK, MW, SvK, KW, NvdK, MS, XvJ and CD were part of the writing committee of the manuscript. All authors reviewed, commented on, and approved the final manuscript.

PESaM Collaborating Group (alphabetical order): Paul Bresser, MD, PhD, Department of Respiratory Medicine OLVG, Amsterdam, The Netherlands. Ilona Coenen, MSc, Department of Pulmonology, Jeroen Bosch Hospital, Hertogenbosch, The Netherlands. Nynke G. Dijkstra, MSc, Department of Respiratory Medicine OLVG,

Amsterdam, The Netherlands. Brigitte A.A. van Dooren, Department of Pulmonology, Jeroen Bosch Hospital, Hertogenbosch, The Netherlands. Miranda J.J. Geelhoed, MD, PhD, Department of Respiratory Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. Marielle Gelens, MD, PhD, Department of Internal Medicine, Division of Nephrology, Maastricht University Medical Centre, Maastricht, The Netherlands. Kirsten de Haas, Department of Respiratory Medicine, Canisius Wilhelmina Ziekenhuis, Nijmegen, The Netherlands. René E. Jonkers, MD, PhD, Department of Respiratory Medicine, AMC, Amsterdam, The Netherlands. Bart Koopman, MD, Department of Respiratory Medicine OLVG, Amsterdam, the Netherlands. Henk Kramer, MD, PhD, Department of Pulmonary Diseases, Martini Hospital, Groningen, the Netherlands. Mirjam J.G. van Manen, PhD, Department of Respiratory Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. Jelle R. Miedema, MD, Department of Pulmonary Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. Karen C.C. Moor, MD, Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands. Rémy L.M. Mostard, MD, PhD, Department of Respiratory Medicine, Zuyderland MC, Heerlen, The Netherlands. Esther J. Nossent, MD, Department of Pulmonary Diseases, VU University Medical Center, Amsterdam, The Netherlands. Marieke J. Overbeek, MD, PhD, Department of Respiratory Medicine, Haaglanden Medical Centre, The Netherlands. Ellen Peeters, Department of Respiratory Medicine, Zuyderland MC, Heerlen, the Netherlands. Rein H.N.A.J. van Rijswijk, MD, Jeroen Bosch Hospital, Hertogenbosch, The Netherlands. Helen Ryan, Department of Pulmonary Diseases, VU University Medical Center, Amsterdam, The Netherlands. Nelleke C. Tak, MSc, Department of Respiratory Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. Anneke van Veen, MD, PhD, Department of Respiratory Medicine, Canisius Wilhelmina Hospital, Nijmegen, The Netherlands. Lucyl Verhoeven, MNAP, Department of Respiratory Medicine, Jeroen Bosch Hospital, Hertogenbosch, The Netherlands. Michiel G.J. de Vries, MD, Department of Pulmonary and Respiratory Medicine, Zuyderland MC, Heerlen, The Netherlands. Monique Wapenaar, MSc, Department of Respiratory Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands. Ingrid M.H. Wegman, MSc, Department of Respiratory Medicine, OLVG, Amsterdam, The Netherlands. Jack Wetzels, MD, PhD, Department of Internal Medicine, Division of Nephrology, Radboud University Medical Center, Nijmegen, the Netherlands.

**Funding** This project was funded by The Federation of Patients and Consumer Organizations in the Netherlands (NPCF), Lung Foundation Netherlands, and The Netherlands Organization for Health Research and Development (ZonMw, Project Number 15200095001).

## Compliance with ethical standards

**Conflict of Interest** Merel Kimman, Marlies Wijnsbeek, Sander van Kuijk, Kioa Wijnsma, Nicole van de Kar, Marjolein Storm, Xana van Jaarsveld, Carmen Dirksen, Paul Bresser, Ilona Coenen, Nynke Dijkstra, Brigitte van Dooren, Miranda Geelhoed, Marielle Gelens, Kirsten de Haas, René Jonkers, Bart Koopman, Henk Kramer, Mirjam van Manen, Jelle Miedema, Karen Moor, Rémy Mostard, Esther Nossent, Marieke Overbeek, Ellen Peeters, Rein van Rijswijk, Helen Ryan, Nelleke Tak, Anneke van Veen, Lucyl Verhoeven, Michiel de Vries, Monique Wapenaar, Ingrid Wegman and Jack Wetzels have no conflicts of interest, including non-financial, that are directly relevant to the contents of this manuscript.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any

noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bensing J. Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. *Patient Educ Couns*. 2000;39(1):17–25.
- Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167.
- Britten N, Pope C, Halford S, Richeldi L. What if we made stratified medicine work for patients? *Lancet Respir Med*. 2016;4:8–10.
- Hirji I, Gupta S, Goren A, Chirovsky DR, Moadel AB, Olavarria E, et al. Chronic myeloid leukemia (CML): association of treatment satisfaction, negative medication experience and treatment restrictions with health outcomes, from the patient's perspective. *Health Qual Life Outcomes*. 2013;11:167.
- Hirsh AT, Atchison JW, Berger JJ, Waxenberg LB, Lafayette-Lucey A, Bulcourn BB, et al. Patient satisfaction with treatment for chronic pain: predictors and relationship to compliance. *Clin J Pain*. 2005;21(4):302–10.
- Mahajan ST, Elkadry EA, Kenton KS, Shott S, Brubaker L. Patient-centered surgical outcomes: the impact of goal achievement and urge incontinence on patient satisfaction one year after surgery. *Am J Obstet Gynecol*. 2006;194:722–8.
- Shingler SL, Bennett BM, Cramer JA, Towse A, Twelves C, Lloyd AJ. Treatment preference, adherence and outcomes in patients with cancer: literature review and development of a theoretical model. *Curr Med Res Opin*. 2014;30(11):2329–41.
- Trawinska MA, Rupesinghe RD, Hart SP. Patient considerations and drug selection in the treatment of idiopathic pulmonary fibrosis. *Ther Clin Risk Manag*. 2016;12:563–74.
- Wijsenbeek M, van Manen M, Bonella F. New insights on patient-reported outcome measures in idiopathic pulmonary fibrosis: only PROMises? *Curr Opin Pulm Med*. 2016;22(5):434–41.
- Wouters H, van Geffen EC, Baas-Thijssen MC, Krol-Warmerdam EM, Stiggelbout AM, Belitser S, et al. Disentangling breast cancer patients' perceptions and experiences with regard to endocrine therapy: nature and relevance for non-adherence. *Breast*. 2013;22(5):661–6.
- Laba TL, Essue B, Kimman M, Jan S. Understanding patient preferences in medication nonadherence: a review of stated preference data. *Patient Patient Cent Outcomes Res*. 2015;8(5):385–95.
- Jin J, Sklar GE, Min Sen OhV, Chuen Li S. Factors affecting therapeutic compliance: a review from the patient's perspective. *Ther Clin Risk Manag*. 2008;4(1):269–86.
- DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care*. 2002;40(9):794–811.
- Golay A. Pharmacoeconomic aspects of poor adherence: can better adherence reduce healthcare costs? *J Med Econ*. 2011;14(5):594–608.
- Facey K, Boivin A, Gracia J, Hansen HP, Lo Scalzo A, Mossman J, et al. Patients' perspectives in health technology assessment: a route to robust evidence and fair deliberation. *Int J Technol Assess Health Care*. 2010;26(3):334–40.
- Hailey D, Werko S, Bakri R, Cameron A, Gohlen B, Myles S, et al. Involvement of consumers in health technology assessment activities by Inahta agencies. *Int J Technol Assess Health Care*. 2013;29(1):79–83.
- Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. *JAMA*. 2008;300(4):436–8.
- Utens CM, Joore MA, van der Weijden T, Dirksen CD. Towards integration of research evidence on patient preferences in coverage decisions and clinical practice guidelines: a proposal for a taxonomy of preference-related terms. *Value Health*. 2014;17(7):583–4.
- Dirksen CD. The use of research evidence on patient preferences in health care decision-making: issues, controversies and moving forward. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(6):785–94.
- van de Bovenkamp HM, Zuiderent-Jerak T. An empirical study of patient participation in guideline development: exploring the potential for articulating patient knowledge in evidence-based epistemic settings. *Health Expect*. 2015;18(5):942–55.
- Kimman ML, Rotteveel AH, Wijsenbeek M, Mostard R, Tak NC, van Jaarsveld X, et al. Development and pretesting of a questionnaire to assess patient experiences and satisfaction with medications (PESaM questionnaire). *Patient Patient Cent Outcomes Res*. 2017;10(5):629–42.
- Boon W, Martins L, Koopmanschap M. Governance of conditional reimbursement practices in The Netherlands. *Health Policy*. 2015;119(2):180–5.
- Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2011;183(4):431–40.
- King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370(22):2083–92.
- Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014;370(22):2071–82.
- Loirat C, Fremeaux-Bacchi V. Atypical hemolytic uremic syndrome. *Orphanet J Rare Dis*. 2011;6:60.
- Verhave JC, Wetzels JF, van de Kar NC. Novel aspects of atypical haemolytic uraemic syndrome and the role of eculizumab. *Nephrol Dial Transpl*. 2014;29(Suppl 4):iv131–41.
- Wijnsma KL, Duineveld C, Volokhina EB, van den Heuvel LP, van de Kar N, Wetzels JFM. Safety and effectiveness of restrictive eculizumab treatment in atypical haemolytic uremic syndrome. *Nephrol Dial Transpl*. 2017. <https://doi.org/10.1093/ndt/gfx196>.
- Strasser S, Aharony L, Greenberger D. The patient satisfaction process: moving toward a comprehensive model. *Med Care Rev*. 1993;50(2):219–48.
- Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
- EuroQol G. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.
- McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1–55.
- Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res*. 1981;18:39–50.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–49.



37. Atkinson MJ, Sinha A, Hass SL, Colman SS, Kumar RN, Brod M, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health Qual Life Outcomes*. 2004;2:12.
38. Al-Jabi SW, Zyoud SH, Sweileh WM, Wildali AH, Saleem HM, Aysa HA, et al. Relationship of treatment satisfaction to health-related quality of life: findings from a cross-sectional survey among hypertensive patients in Palestine. *Health Expect*. 2015;18(6):3336–48.
39. Khanna PP, Shiozawa A, Walker V, Bancroft T, Essoi B, Akhras KS, et al. Health-related quality of life and treatment satisfaction in patients with gout: results from a cross-sectional study in a managed care setting. *Patient Prefer Adher*. 2015;9:971–81.
40. Hinkle DE, Wiersma W, Jurs SG. *Applied statistics for the behavioral sciences*. 5th ed. Boston: Houghton Mifflin; 2003.
41. Bland JM. *An introduction into medical statistics*. Oxford: Oxford University Press; 2000.
42. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48(2):1–36.
43. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22(1):85–93.
44. Ruiz MA, Pardo A, Rejas J, Soto J, Villasante F, Aranguren JL. Development and validation of the “Treatment Satisfaction with Medicines Questionnaire” (SATMED-Q). *Value Health*. 2008;11(5):913–26.

## Affiliations

**Merel L. Kimman<sup>1</sup> · Marlies S. Wijsenbeek<sup>2</sup> · Sander M. J. van Kuijk<sup>1</sup> · Kioa L. Wijnsma<sup>3</sup> · Nicole C. A. J. van de Kar<sup>3</sup> · Marjolein Storm<sup>4</sup> · Xana van Jaarsveld<sup>5</sup> · Carmen D. Dirksen<sup>6</sup> · PESaM Collaborating Group**

Marlies S. Wijsenbeek  
m.wijsenbeek-lourens@erasmusmc.nl

Sander M. J. van Kuijk  
sander.van.kuijk@mumc.nl

Kioa L. Wijnsma  
Kioa.Wijnsma@radboudumc.nl

Nicole C. A. J. van de Kar  
Nicole.vandeKar@radboudumc.nl

Marjolein Storm  
storm@nvn.nl

Xana van Jaarsveld  
XanavanJaarsveld@longfonds.nl

Carmen D. Dirksen  
c.dirksen@mumc.nl

<sup>1</sup> Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre, Maastricht, The Netherlands

<sup>2</sup> Department of Respiratory Medicine, Erasmus Medical Centre, University Hospital Rotterdam, Rotterdam, The Netherlands

<sup>3</sup> Department of Pediatric Nephrology, Radboud University Medical Center, Radboud Institute for Health Sciences, Amalia Children’s Hospital, Nijmegen, The Netherlands

<sup>4</sup> Kidney Patients Association, Bussum, The Netherlands

<sup>5</sup> Lung Foundation Netherlands, Amersfoort, The Netherlands

<sup>6</sup> Department of Clinical Epidemiology and Medical Technology Assessment, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Centre, Maastricht, The Netherlands