Volume 5 • Number 5 • 2002
VALUE IN HEALTH

Validation of the Patient Perception of Migraine Questionnaire

Kimberly Hunt Davis, MS,[1] Libby Black, PharmD,[1] Betsy Sleath, PhD[2]

[1]GlaxoSmithKline, Research Triangle Park, NC, USA; [2]University of North Carolina, Chapel Hill, NC, USA

ABSTRACT

Objective: The purpose of this study was to assess the psychometric properties of the Patient Perception of Migraine Questionnaire (PPMQ), which measures patient satisfaction with migraine therapy.

Methods and Data: The PPMQ was administered to 940 patients as part of a 3-month, multinational, open-label, clinical trial comparing the effects of oral naratriptan 2.5 mg with the patient's customary therapy for the treatment of migraine. Psychometric properties of the PPMQ were evaluated in terms of its latent factor structure, validity, reliability, sensitivity, and development of a scoring method. Classical Test theory and Item Response theory (IRT) modeling were both used to measure reliability.

Results: The PPMQ was able to detect treatment differences ($P > .001$), and all items significantly correlated with diary ratings of headache pain ($r = .18–.51$, $p > .0001$) and the Medical Outcomes Short Form-36 pain scale ($r = .27$, $p > .0001$). A principal components factor analysis revealed that the items on the PPMQ were psychometrically distinct and unidimensional (loadings, 0.74–0.91), with the exclusion of two items. The reliability (i.e., internal item consistency) of the PPMQ post-trial was high in both treatment groups (Cronbach's $\alpha = 0.96$). An IRT analysis also ensured the formation of homogenous items, which were stable on repeat administration. Items did not require weighting and can be simply summed to yield a total score.

Conclusion: Based on the data from this one clinical trial, the 15-item PPMQ was shown to be a valid and reliable instrument that seems to efficiently and comprehensively measure patient perception of drug attributes in relation to the treatment of symptoms associated with migraine headaches.

Keywords: Patient Perception of Migraine Questionnaire, psychometric validation, migraine, satisfaction, naratriptan.

Introduction

While some individuals may not have the requisite clinical knowledge to judge the quality of drug therapy, nearly all patients have expectations about the medication that their physician prescribes for them. If the experience meets or exceeds that expectation, then the patient is usually satisfied. Conversely, if the experience does not match the expectation, then the patient is likely to be dissatisfied. Dissatisfied patients in the United States are often noncompliant, often delay seeking medical care, are quick to quit medical insurance plans, and often lack continuity of care due to "doctor shopping" or continuous switching of their primary physician [1]. Thus, patient satisfaction has become an important benchmark to gauge the quality of health care.

Although satisfaction with health-care delivery and patient–provider relationships has been extensively studied, there are few studies that directly measure patient satisfaction with drug therapy. Pharmaceutical development would likely benefit from an effort to develop questionnaires that effectively measure satisfaction with treatment, because patient satisfaction with a medication may play an important role in the selection, utilization, and ultimate effectiveness of that treatment. To be used appropriately, however, such questionnaires must first be standardized, which requires assessment of their validity and reliability.

Address correspondence to: Kimberly Hunt Davis, MS, Global Health Outcomes, GlaxoSmithKline, PO Box 13398, Research Triangle Park, NC 27709-3398.
E-mail: khd64468@gsk.com

 1098-3015/02/$15.00/422 422–430

Research shows that how well a drug works, how safe it is, how fast it works, and side effects are important key attributes of drug therapy for the treatment of migraine headaches [2]. An additional study concluded that complete pain relief, no migraine recurrence, rapid onset, no side effects, and relief of associated symptoms are important attributes for the acute treatment of migraine attacks [3]. We used this information in the development of the Patient Perception of Migraine Questionnaire (PPMQ), which was designed to measure patient satisfaction with migraine therapy. This study evaluated the PPMQ's psychometric properties: latent factor structure, validity, reliability, and sensitivity. The PPMQ was administered as part of a multinational clinical study comparing the effects of the new migraine treatment naratriptan with that of the patient's customary migraine therapy. If satisfaction with drug therapy can be determined reliably with the PPMQ, it is hoped that this may then translate into earlier and closer to optimal clinical outcomes for patients with migraine, with secondary benefits to both the health-care provider and the insurers.

Methods

Study Design

Data were collected as part of an open-label, randomized, parallel-group clinical trial comparing 3 months of treatment with oral naratriptan 2.5 mg (n = 481) with the patient's customary therapy (n = 474) in six countries: Spain, New Zealand, the Netherlands, Hungary, Finland, and Canada [4]. Treatment for the control group included one or more customary therapies, excluding any 5-$HT_{1B/1D}$ receptor agonists. Subjects received a diagnosis of migraine according to the International Headache Society criteria. Patients rated the level of pain from onset of each migraine to relief on a 4-point pain scale (0 = none; 1 = mild; 2 = moderate; 3 = severe) and reported their pain assessment on a diary card. The primary measure of efficacy was the relief of headache, defined as a reduction of patient-rated pain from moderate or severe to mild or none, 4 hours after dosing with naratriptan. A widely used, general health-related quality of life questionnaire, the Medical Outcomes Short Form-36 (SF-36) [5], was administered in addition to the PPMQ at baseline and at 3 months post-trial.

PPMQ

The PPMQ was based on an earlier questionnaire (Migraine Treatment Questionnaire), which was developed to assess the impact of migraine and migraine pharmacotherapy on patients. This original questionnaire was pretested in a sample of 24 patients with migraine who were participating in a clinical trial. Based on participants' comments, further modifications were made to the questionnaire, resulting in a 7-item questionnaire with a 5-point Likert-type response scale. Analysis after the piloting of this 7-item questionnaire in a clinical trial indicated that further refinements to the questionnaire were necessary. A series of three focus groups (10–15 patients in each) and 15 patient interviews resulted in several modifications of the questionnaire including addition of 8 new items, changes to existing items, and a change from a 5- to a 7-point response scale. During the focus groups and patient interviews, patients were asked to rank each of the items in terms of importance. Those attributes deemed to be of greatest importance (i.e., at least 75% of patients ranked the attribute as important to very important) were considered the "core" questions (items 1a–2a), and the remaining 7 items were considered optional or could be incorporated into studies as needed.

The present trial was conducted in six countries, and the 15-item PPMQ was translated from the original English version forward and backward into the appropriate languages and was administered at baseline and post-trial [6]. Baseline instructions stated that patients should respond to questions in the PPMQ based on their experience with previous therapy. At the protocol scheduled 3-month post-trial visit, patient responses were based on treatment for migraine headache administered within the previous 3 months. The first 8 items of the PPMQ (Appendix A) were recorded on a 7-point summated rating Likert-type scale (very dissatisfied to very satisfied), and responses for the remaining 7 items were recorded on customized 5-point summated Likert-type rating scales.

Data Analysis

Only subjects completing both study visits were included in the analyses that depended on the data from these visits. All statistical tests were conducted in the intent-to-treat population using "last observation carried forward" methodology, which accounts for dropouts in both treatment arms with alpha set at 0.05, and two-tailed *P* values are reported. A descriptive analysis was performed initially to evaluate the mean scores of individual items, the standard deviation, the variance, frequency of responses, and a summary of missing data at baseline and at post-trial. The data were

then evaluated for latent factor structure, validity, reliability, and sensitivity. Because a problem in translation existed in Hungary for item 2g of the PPMQ, data from Hungary were excluded from all analyses.

The latent factor structure of the PPMQ was evaluated using a principal components varimax rotated and oblique rotated analysis using SAS (version 6.12, 1990; SAS Institute, Inc., Cary, NC, USA). To determine the number of factors within a scale, Kaiser's eigenvalue rule [7] was used, which states that any factor with an eigenvalue greater than 1.0 should be considered important when accounting for item variability. Item loadings of 0.3 and above are considered significant [8].

To assess validity of the PPMQ, or its ability to discriminate between the two treatment groups, treatment comparisons were performed using a two-tailed Students' *t* test of the mean differences in baseline and post-trial PPMQ scores. Validity of the individual items was determined using analysis of covariance, adjusted for country, of the mean change from baseline to the end of the 3-month treatment period for each item in each treatment group. Additionally, a subgroup analysis of patients—those who experienced less than six migraine attacks and those who experienced more than six migraine attacks—was performed. Validity was also measured with the Pearson correlation coefficients between each item of the PPMQ and pain scores from diary cards, between the total score of the PPMQ and pain scores from diary cards, and between selected items on the PPMQ and SF-36.

To measure the reliability of the PPMQ, Classical Test theory (CTT) and Item Response theory (IRT) modeling were used. IRT is a nonlinear probabilistic model that specifies a relationship between observable patient response and the patient's underlying feelings, as reflected by a questionnaire score [9]. The purpose of our IRT analysis was to model the interaction between the choice of a patient's item response and the patient's level of satisfaction. Thus, the standard error is dependent and based on each patient's perception. In CTT, the standard error is a constant value determined by the distribution of the sample and is assumed to be the same for everyone in the sample; individual differences among items or subjects are treated as random error [7]. Item internal consistency in the CTT analysis was considered satisfactory if the correlation between an item and its hypothesized scale was at least 0.40 [10]. Reliability of the PPMQ was evaluated using Cronbach's alpha coefficient, which yielded a coefficient of ≥ 0.7 [7], evidence of the instrument's reliability.

Three parameters (theta, alpha, beta) were used in the IRT analysis, using MULTILOG software (Version 6.0, 1991; Scientific Software International, Inc., Lincolnwood, IL, USA), to examine the distribution of responses and the response characteristics of individual items or internal consistency reliability [9]. The scale of latent satisfaction is defined by setting the mean of the distribution of the examined treatment groups to 0, and the standard deviation is set to 1 for standard IRT analysis. When comparing parameters from two different groups, the parameters of both groups must be measured on the same scale. In our model, theta represents the subject's level of satisfaction, predicting the probability of choosing a given response for a specific item. Alpha represents the degree of discrimination that the item provides between persons at different levels of satisfaction or theta. Items for which alpha is less than 0.7 are rarely of any practical value and are not considered to be measuring satisfaction [11]. The beta parameter represents a subject's threshold of satisfaction or the point at which the subject's experience with the drug therapy meets the subject's expectation. The number of threshold parameters is the number of response categories minus 1, because the thresholds define the point at which subjects are equally likely to choose adjacent response items. Item thresholds need to be well distributed to distinguish between lower and higher levels of satisfaction. Beta parameter estimates typically range from about –2.0 (very dissatisfied) to +2.0 (very satisfied) [9].

Sensitivity, or responsiveness, of our instrument was measured in two ways: the difference between baseline and post-trial mean total PPMQ scores was calculated for each country to assess whether the instrument could discriminate within treatment groups, and the difference between pre- and post-trial pain severity scores from patient diaries as described above was calculated for each migraine attack and the mean calculated for each individual over the 3-month treatment period. These mean differences in pain scores were then plotted against the total post-trial PPMQ scores for each patient.

Results

The mean age of the overall sample (N = 793) at baseline was 38.4 years; 85% were female and 98% were white, which reflects multinational prevalence data [12]. The number of patients in each treatment group was approximately the same in each country,

Table 1 Patient demographics at baseline

Country	Characteristic	Naratriptan	Customary therapy
Canada	No. of patients	114	110
	Mean age, years (SD)	40.2 (8.7)	39.0 (9.9)
	Women (%)	88	86
Finland	No. of patients	98	98
	Mean age, years (SD)	36.7 (10.7)	36.8 (10.8)
	Women (%)	86	87
The Netherlands	No. of patients	65	62
	Mean age, years (SD)	37.9 (9.6)	41.7 (12.1)
	Women	88	82
New Zealand	No. of patients	59	55
	Mean age, years (SD)	39.6 (10.6)	41.3 (10.0)
	Women (%)	73	75
Spain	No. of patients	61	64
	Mean age, years (SD)	36.2 (9.1)	33.0 (9.2)
	Women (%)	84	81

and treatment groups had similar sex and age distributions (Table 1). At baseline, the average length of time that a subject had experienced migraines was 18.1 years in the naratriptan group and 17.4 years in the customary therapy group. Over the 3-month clinical trial period, the average number of migraine attacks was similar in the two treatment groups within each country. A summary report of acute migraine medication use before entering the study indicated that acetaminophen (26%), ibuprofen (20%), other nonsteroidal drugs (26%), and aspirin (6%) were the most commonly used medications in each treatment group. Customary therapy during the trial was most frequently nonsteroidal anti-inflammatory drugs (43%), analgesics (24%), and ergot alkaloids (11%).

The variances for each item on the PPMQ were approximately equal at baseline and post-trial, and the distribution of responses for each item was only slightly skewed to the right at baseline, with minimal ceiling (<13%) and floor (<10%) effects. Because the data were normally distributed, the data for each country were pooled for the analyses.

Latent Factor Structure

The initial principal components analysis with a varimax rotation indicating a two-factor solution is summarized in Table 2. Factors 1 and 2 explained 62% and 11% of the total variance, respectively. All of the items loaded relatively high on the factor 1 solution, except item 1d ("how drowsy the drug makes you") and item 2g ("how easy the drug was to use") for the sample of all countries combined as well for each individual country, except Spain. An oblique rotation also yielded a similar solution as the varimax rotation. The interfactor correlation for the oblique rotation was 0.39. Loadings on factor 2 were high with both rotations for only items 1d and 2g, but face validity suggests that these items do not appear to be measuring a common latent variable. For Spain, item 2f (prevents recurrence) loaded on factor 2 in addition to item 1d and 2 g.

Validity

The PPMQ demonstrated construct validity in that it was able to detect treatment differences at the end of the study when data from all countries were combined into one sample. The change in mean PPMQ total scores from baseline to post-trial was statistically significant in the naratriptan group ($P = .0001$) but not in the customary therapy group ($P = .884$), indicating greater satisfaction with naratriptan than with customary therapy. As summarized in Table 3, changes in mean scores from baseline to post-trial of each individual item were also significantly different ($P < .001$) when the naratriptan group was compared to the group receiving customary therapy. Some subjects may not have treated a sufficient

Table 2 Latent factor structure: principal components with varimax rotation (N = 650)

Item	Factor 1	Factor 2
Pain relief (1a)	0.907	0.171
Other symptoms relieved (1b)	0.818	0.190
Speed of relief (1c)	0.873	0.199
Drowsiness (1d)	0.276	0.678
Length of time it works (1e)	0.852	0.244
No. of doses (1f)	0.866	0.202
Return to activities (1g)	0.814	0.312
Effective overall (1h)	0.914	0.221
How fast (2a)	0.839	0.196
Resume to activities (2b)	0.796	0.272
How consistent (2c)	0.830	0.167
How completely (2d)	0.886	0.164
How long (2e)	0.804	0.118
Prevents recurrence (2f)	0.737	0.086
Ease of use (2g)	0.067	0.836

Table 3 Construct validity: detection of treatment differences*

Item	Mean change in score		Difference between groups† (SE)
	Naratriptan	Customary therapy	
Pain relief (1a)	1.08	–0.10	1.18 (0.11)
Other symptoms relieved (1b)	1.27	0.00	1.27 (0.11)
Speed of relief (1c)	1.06	–0.11	1.17 (0.12)
Drowsiness (1d)	0.73	0.06	0.66 (0.10)
Length of time it works (1e)	1.27	0.10	1.17 (0.11)
No. of doses (1f)	1.51	0.07	1.44 (0.12)
Return to activities (1g)	1.43	0.14	1.29 (0.12)
Effective overall (1h)	1.42	0.04	1.38 (0.12)
How fast (2a)	0.70	0.05	0.65 (0.07)
Resume to activities (2b)	0.76	0.09	0.67 (0.07)
How consistent (2c)	0.75	–0.01	0.76 (0.07)
How completely (2d)	0.90	0.07	0.83 (0.07)
How long (2e)	0.78	0.09	0.68 (0.07)
Prevents recurrence (2f)	0.79	0.18	0.60 (0.07)
Ease of use (2g)	0.37	0.05	0.33 (0.05)

*Mean change in scores from baseline to post-trial by PPMQ item.
†$P < .001$, difference between naratriptan and customary therapy for each item.

number of attacks over the 3 months to allow treatment differences to manifest; however, similar results (i.e., significantly greater satisfaction with naratriptan than with customary therapy) were observed in the subanalysis of data from subjects who experienced either less than six migraine attacks or more than six migraine attacks.

Average pain relief scores recorded on diary cards 4 hours post-treatment during the study were significantly correlated ($P < .0001$, $r = .18–.51$) with post-treatment responses to each PPMQ item as well as with the total score as per Table 4, another indication of the construct validity of the PPMQ. Scores on the two pain-related SF-36 questions ("How much bodily pain have you had during the past 4 weeks?" and "During the past 4 weeks, how much did pain interfere with your normal work?") were significantly correlated ($P \leq .002$, $r = .14 – .27$) with PPMQ pain items 1a (how well the medication relieves pain), 1g (quickness of return to usual activities after using the medication), and 2b (how fast it allows return to usual day-to-day activities).

Reliability

Using the CTT, all items on the PPMQ met the criteria for internal consistency (item-scale: $r > .4$), except items 1d ($r = .36$) and 2g ($r = .13$) as indicated in Table 5. At baseline, the PPMQ (with and without items 1d and 2g) for all countries combined resulted in a Cronbach's alpha coefficient of 0.94. The post-trial reliability estimate was 0.96 for both

Table 4 Construct validity: correlation of average pain relief with PPMQ items

PPMQ item	Pearson correlation*	N
Pain relief (1a)	.50	674
Other symptoms relieved (1b)	.46	671
Speed of relief (1c)	.48	674
Drowsiness (1d)	.30	662
Length of time it works (1e)	.46	666
No. of doses (1f)	.51	669
Return to activities (1g)	.49	674
Effective overall (1h)	.49	665
How fast (2a)	.44	656
Resume to activities (2b)	.46	655
How consistent (2c)	.40	656
How completely (2d)	.45	655
How long (2e)	.40	655
Prevents recurrence (2f)	.35	656
Ease of use (2g)	.18	657
Total score	.67	632

*$P = .0001$ for each PPMQ item.
PPMQ, Patient Perception of Migraine Questionnaire.

Table 5 Reliability: item-scale (Pearson) correlations*

Item	Baseline	Post-trial
Pain relief (1a)	.83	.91
Other symptoms relieved (1b)	.72	.81
Speed of relief (1c)	.84	.88
Drowsiness (1d)	.36	.47
Length of time it works (1e)	.80	.87
No. of doses (1f)	.77	.87
Return to activities (1g)	.80	.86
Effective overall (1h)	.87	.92
How fast (2a)	.78	.84
Resume to activities (2b)	.77	.81
How consistent (2c)	.72	.79
How completely (2d)	.80	.87
How long (2e)	.68	.78
Prevents recurrence (2f)	.57	.65
Ease of use (2g)	.13	.46

*For internal consistency, items should demonstrate a correlation coefficient of $r > .4$.

Table 6 Reliability: IRT analysis of alpha and beta parameter estimates at baseline and post-trial*

	Baseline item parameters							Post-trial item parameters						
Item	α	β1	β2	β3	β4	β5	β6	α	β1	β2	β3	β4	β5	β6
1a	4.24	−1.37	−0.45	0.39	0.55	0.95	1.77	3.34	−1.9	−0.75	0.31	0.62	1.17	2.09
1b	2.41	−1.97	−0.73	−0.08	0.49	0.94	1.85	2.00	−2.25	−0.83	−0.06	0.68	1.22	2
1c	4.21	−1.50	−0.71	0.07	0.26	0.78	1.58	2.62	−2.32	−1.08	−0.2	0.22	0.93	1.98
1d	0.72	−3.36	−1.02	−0.03	2.18	3.66	5.28	0.6	−3.09	−0.79	0.01	3.11	4.62	6.18
1e	3.14	−1.75	−0.62	0.02	0.29	0.83	1.62	2.35	−2.14	−0.82	−0.07	0.54	1.27	2.15
1f	2.70	−1.83	−0.77	−0.21	0.14	0.72	1.61	2.48	−1.97	−0.84	−0.15	0.38	0.95	1.95
1g	3.25	−1.78	−0.80	−0.14	0.09	0.72	1.47	2.16	−2.29	−0.96	−0.09	0.34	1.08	2.01
1h	5.75	−1.59	−0.59	0.05	0.21	0.62	1.28	3.93	−1.99	−0.84	0	0.37	0.92	1.72
2a	3.20	−1.86	−0.79	0.41	1.19			2.20	−2.73	−1.01	0.5	1.75		
2b	2.66	−2.09	−1.09	0.26	1.23			1.88	−2.94	−1.1	0.54	1.91		
2c	2.20	−2.02	−0.61	0.37	1.60			1.96	−2.48	−0.76	0.38	1.76		
2d	2.96	−1.95	−0.85	0.35	1.47			2.55	−2.28	−0.77	0.47	1.85		
2e	2.03	−2.16	−0.97	0.71	1.71			1.65	−2.47	−0.85	0.95	2.08		
2f	1.36	−2.79	−1.11	0.04	1.47			1.28	−3.00	−1.02	0.40	1.83		
2g	0.24	0.45	8.29	12.89	18.29			0.43	0.98	6.53	10.09	11.72		

*α, degree of discrimination that the item provides between persons at different levels of satisfaction (α must be > 0.7 to discriminate). β, the ability of each item to capture subjects with thresholds of very high or very low satisfaction (i.e., point at which the subject's experience with the drug equals expectation).
Abbreviation: IRT, Item Response theory.

treatment groups and was consistent with or without items 1d and 2g providing additional support for the reliability of the PPMQ. Cronbach's alpha at post-trial was also very similar within each country: Canada (0.96), Finland (0.96), the Netherlands (0.95), New Zealand (0.96), and Spain (0.95).

Internal consistency reliability was also confirmed by IRT analysis, with a marginal reliability of 0.95 at baseline and post-trial. All of the alpha parameters were greater than 1.0 at baseline and post-trial (Table 6), except items 1d and 2g, indicating the ability of each item to discriminate. Most of the items other than 1d and 2g had a broad range of beta parameter estimates indicating the ability of each item to capture subjects with thresholds of very high or very low satisfaction. The standard errors ranged from 0.05 to 0.19 for each beta item parameter estimate excluding items 1d and 2g, indicating precision and item effectiveness.

Sensitivity

The PPMQ was able to discriminate between treatment groups in each of the countries. The difference between baseline and post-trial mean total scores was not significant ($P > .1$) for the customary therapy group in any country, while they were significant ($P \leq 0.004$) for the naratriptan group in each country as summarized in Table 7. When the mean change in pain-severity scores (from pre-treatment to post-trial) was graphed against mean post-trial total PPMQ scores, the PPMQ scores increased (improved) as the difference in severity rating increased (improved). Differences in these severity scores ranged from –3.0 to 3.0, with a score above zero representing an improvement in migraine at 4 hours post-trial. Sensitivity of the PPMQ is represented by the positive linear slope in Figure 1.

Table 7 Sensitivity: mean differences in baseline and post-trial scores

Treatment/country	Mean baseline score (SE)	Mean change from baseline (SE)	Student's *t* test	P > |T|
Naratriptan				
Canada (n = 96)	48.85 (16.92)	9.86 (26.26)	3.68	.0004
Finland (n = 82)	48.70 (13.56)	10.04 (21.84)	4.16	.0001
The Netherlands (n = 49)	55.77 (15.55)	16.96 (20.84)	5.70	.0001
New Zealand (n = 44)	51.64 (13.62)	18.50 (20.54)	6.01	.0001
Spain (n = 47)	47.0 (16.37)	12.42 (19.53)	4.36	.0001
Customary therapy				
Canada (n = 91)	49.49 (14.86)	1.66(15.53)	1.02	.31
Finland (n = 88)	49.43 (15.75)	1.42 (16.23)	0.82	.41
The Netherlands (n = 43)	53.27 (16.16)	–2.60 (13.82)	–1.24	.22
New Zealand (n = 40)	57.67 (18.35)	0.20 (18.23)	0.07	.95
Spain (n = 52)	47.41 (16.64)	–4.11 (17.56)	–1.69	.10

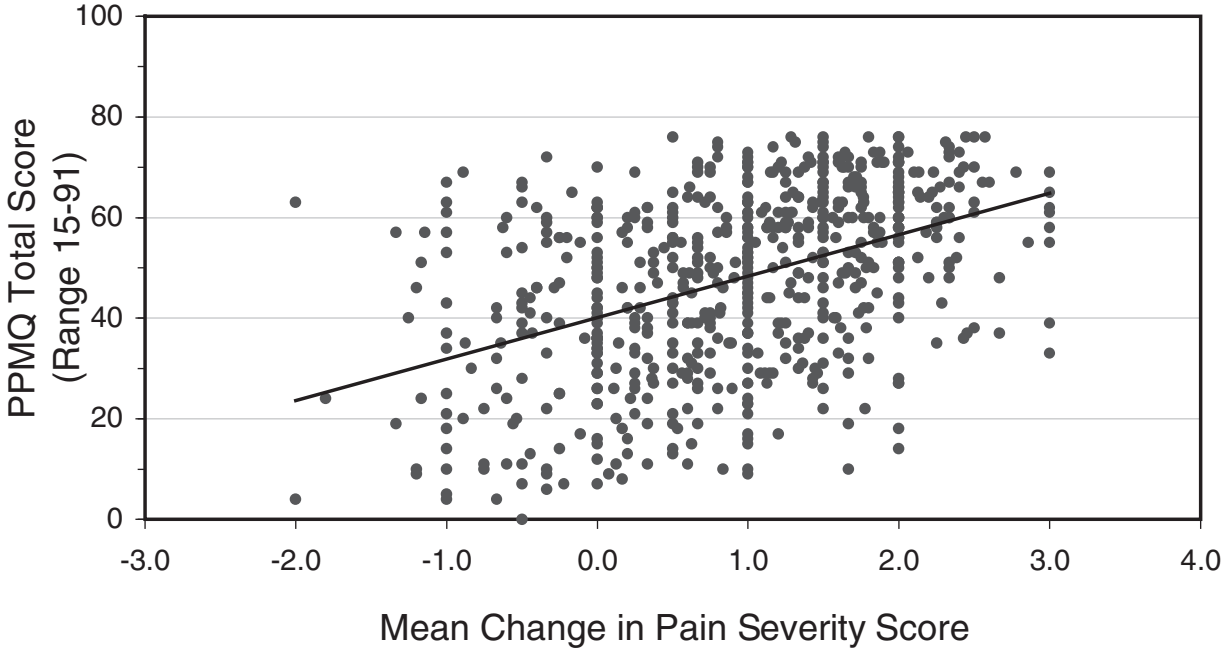

Figure 1 Mean PPMQ total post-trial scores (range, 15–91) versus mean change in pain severity scores from patient diaries.

Scoring

Item-scale correlations ranged between 0.57 and 0.87 with the exception of items 1d (0.36) and 2g (0.13) suggesting that these two items should be excluded from scoring. With the exception of these two items, these results provide strong empirical support that each item contributes roughly equal and substantial proportions of information to the total scale score and therefore no weighting is required. Responses to each item are not weighted and are simply summed to yield a total score with a range of 15 to 91. Higher scores indicate greater satisfaction.

Discussion

This study demonstrates that the PPMQ, with some refinements to the questionnaire, measures patient satisfaction with migraine treatment both reliably and validly. Internal consistency reliability for the PPMQ (0.96) far exceeded the standard of 0.70 generally applied to self-reported instruments used for group assessment of health-related quality of life [7]. For an individual assessment, the standard for internal consistency reliability is 0.9, which would allow physicians to assess individual patients in terms of patient preference and satisfaction for a given migraine therapy in a clinical setting [13,14]. The principal components factor analysis with varimax rotation revealed that the satisfaction subscale was conceptually distinct, with the exception of items 1d and 2g. The questionnaire was able to detect score differences between the two treatment groups. The degree of sensitivity or responsiveness was consistent within each country and consistently measured a minimal mean difference of 9 points in detecting differences between the treatment groups in this study.

Face validity shows that the following pairs of items in the PPMQ are similarly worded: 1c, how fast the medication relieves migraine pain and other migraine symptoms; 2a, how fast it starts to relieve migraine pain; 1a, how well the medication relieves pain; 1h, how effective the medication is overall at relieving migraine pain and other migraine symptoms; 1e, how long the medication works; 2e, how long it relieves migraine pain; 1g, quickness of return to usual activities after using the medication; and 2b, how fast it allows return to usual day-to-day activities. Similarly worded items did not appear sequentially in the questionnaire, however, which may account for local dependence not being violated in the IRT analysis as well as for the high estimates of internal consistency reliability. Because face validity suggests redundancy, and item-scale correlations were high for these item pairs, we would therefore recommend that four items (1h, 2a, 2b, and 2e) be dropped from future versions of the

PPMQ. These four items were chosen because each had a lower alpha (discriminatory) IRT parameter compared to its corresponding similarly worded item.

Moderately significant correlations were observed between the clinical efficacy of naratriptan and satisfaction with this therapy ($r = .46 - .51$), despite the limitations of having to compare a 4-point pain severity scale from patient diaries with the 7-point Likert-type scale in the PPMQ or a generic quality of life scale (SF-36) with a migraine-specific satisfaction scale. The pain items in the SF-36 also reflected only the last 4 weeks of the trial, whereas no time limits were stated in the PPMQ.

The analysis concludes that 13 of the 15 items on the PPMQ constitute a unidimensional item pool, excluding items 1d (how drowsy the medication makes you) and 2g (ease of use). The PPMQ was initially developed to provide a pool of questions, which would cover attributes of any given migraine drug in terms of satisfaction. Despite the overall success of the PPMQ, these two items did not perform well in this study. Drowsiness was not a common side effect of naratriptan (the only triptan available at the time) nor was it for many of the customary therapies seen in this trial, which may explain why this item did not discriminate. Because additional triptans are now available, possibly with different side-effect profiles, a new item will be added to the PPMQ that attempts to assess satisfaction with side effects other than drowsiness, and this version will be piloted in future studies. Item 2g (ease of use) was included on the questionnaire to capture preference for a given dosing regimen. Because the majority of dosage forms for this trial were oral tablets, this item did not discriminate among individuals in terms of a preference for a given formulation. However, item 2g would provide versatility for use in future trials that may include other dosage forms, such as nasal sprays or injectables. Alternatively, both of these items may be important constructs and could be retained as optional items to be reported separately and not scored.

Item 2f (how consistently it prevents my migraine from coming back) could be further refined. Although the item loaded at 0.737 on factor 1 in the principal component factor analysis, the item-scale correlations were 0.57 at baseline and 0.65 at post-trial. Although the item does reflect a marginal improvement at post-trial, the reason for the moderate estimates may be that only a subset of patients experienced recurrence: 26% of attacks treated with naratriptan and 28% of attacks treated with customary therapy. This assertion cannot be empirically tested because the study was a multiattack study and recurrence does not occur with every migraine attack. The word "recurrence" was purposefully not chosen when the PPMQ was developed initially because the term is ambiguous and could be defined differently by different individuals. A possible solution might be to add a response of N/A (not applicable), which would exclude those individuals who do not have recurrent migraines.

Interpretation of the study results may be confounded by the study design, the frequency of migraine attacks, the severity of individual attacks, and recall bias. Although the trial was randomized, subjects could not be blinded because of inclusion of different formulations and drugs in the customary therapy group and acute dosing restrictions for naratriptan. Because this was an open-label trial, bias may have occurred in the customary therapy group, resulting in overestimation of item-scale correlations. No subgroup analyses were performed within the customary therapy group because the sample size was too small for each class of drugs. Consequently, the data did not allow for control of within treatment differences for the customary therapy group.

Recall bias on the PPMQ questionnaire may have occurred because of the 3-month clinical trial time frame. Most subjects in each of the treatment groups reported more than one headache during the trial, and headache severity often varied for individuals. Responses to the PPMQ may reflect satisfaction with drug therapy for the most severe headaches rather than a composite of all of the headaches experienced. However, the bias should have been nondifferential, because it would be expected to have occurred in both treatment groups.

In summary, the results of this five-country study indicate that the PPMQ is both valid and reliable. For future studies, we recommend some minor modifications to the questionnaire, including addition of one item (side effects) and deletion of several redundant items.

The authors thank Rose Mills for assistance in writing the manuscript.

This research was funded in part by Glaxo Wellcome Inc.

References

1 DiTomasso RA, Willard M. The development of a patient satisfaction questionnaire in the ambulatory setting. Fam Med 1991;23:127–31.
2 Luciani RJ, Osterhaus JT, Gutterman DL. Patient preferences for migraine therapy: subcutaneous sumatriptan compared with other medications. J Fam Prac 1995;41:147–52.
3 Lipton RB, Stewart WF. Acute migraine therapy: do doctors understand what patients with migraine want from therapy? Headache 1999; 39(Suppl 2):S20–6.
4 Bayliss MS, Black L, Ware JE. Patient satisfaction with migraine therapy: naratriptan compared with customary therapy. Cephalalgia 2000;20:411.
5 Ware JE. SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center, 1993.
6 Conway K, Uzun V, Black L. Linguistic validation of the Patient Perception of Migraine Questionnaire (PPMQ) in 16 languages. Value Health 1999;2:206.
7 Nunnally JC, Bernstein IH. Psychometric Theory (3rd ed.). New York: McGraw-Hill, 1994.
8 Kline P. An Easy Guide to Factor Analysis. New York: Routledge, 1999.
9 Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage, 1991.
10 Ware JE, Harris WJ, Gandek B, et al. Multitrait/Multi-item Analysis Program—Revised. Boston, MA: Health Assessment Lab, 1997.
11 Thissen D. MULTILOG User's Guide—Multiple, Categorical Item Analysis and Test Scoring Using Item Response Theory (Version 6.0). Chicago: Scientific Software, 1991.
12 Dahlof C, Bouchard J, Cortelli P, et al. A multinational investigation of the impact of subcutaneous sumatriptan. II. Health-related quality of life. Pharmacoeconomics 1997;11(Suppl 1):24–34.
13 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use (2nd ed.). New York: Oxford University Press, 1995.
14 McDowell I, Newell C. Measuring Health: A Guide to Rating Scales and Questionnaires (2nd ed.). New York: Oxford University Press, 1996.

Appendix A

Summarized Version of the PPMQ

1a. How well the medication relieves pain
1b. How well the medication relieves other migraine symptoms
1c. How fast the medication relieves migraine pain and other migraine symptoms
1d. How drowsy the medication makes you feel
1e. How long the medication works
1f. Number of doses needed for relief of symptoms
1g. Quickness of return to usual activities after using the medication
1h. How effective the medication is overall at relieving migraine pain and other migraine symptoms
2a. How fast it starts to relieve migraine pain
2b. How fast it allows return to usual day-to-day activities
2c. How consistently it relieves migraine pain
2d. How completely it relieves migraine pain
2e. How long it relieves migraine pain
2f. How consistently it prevents migraine pain from coming back
2g. How easy it is to use