

The Headache Under-Response to Treatment (HURT) Questionnaire: Assessment of utility in headache specialist care

Maria LS Westergaard¹, Timothy J Steiner^{2,3},
E Anne MacGregor⁴, Fabio Antonaci⁵, Cristina Tassorelli⁶,
Dawn C Buse^{7,8}, Richard B Lipton^{7,8} and Rigmor H Jensen¹

Cephalalgia
33(4) 245–255
© International Headache Society 2012
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0333102412469740
cep.sagepub.com


Abstract

The HURT Questionnaire consists of eight questions which the patient answers as a measure of effectiveness of intervention against headache. This first assessment of clinical utility was conducted in headache specialist centres in three countries in order to demonstrate that HURT was responsive to change induced by effective management. We administered HURT on three occasions to 159 consecutive patients seeking non-urgent care from centres in Denmark and the United Kingdom: the first before the initial visit to the centres; the second at the initial visit; and the third when the specialist judged that the best possible outcome had been achieved in each patient. Questionnaires were also answered by 42 patients at initial and final visits to a centre in Italy. Internal consistency reliability was very good ($\alpha = 0.85$) while test-retest reliability was fair to low ($\kappa = 0.38$ – 0.62 and $r_s = 0.49$ – 0.76), possibly because headache was unstable prior to start of management. There were significant changes in responses post-intervention compared with baseline ($p < 0.01$), indicating a favourable outcome overall in up to 77% of patients, and responsiveness to change, but there was no improvement in patients' concerns about side effects of medication ($p = 0.18$). We conclude that the questionnaire has utility across headache disorders. It can help patients describe headache frequency and headache-attributed disability, medication use/efficacy/tolerability, self-efficacy and knowledge about headache. It may guide physicians in assessment of disability of individual patients, how to proceed with management towards the best possible outcome, and in evaluating the quality of management.

Keywords

Headache disorders, migraine, tension-type headache, outcome measure, Global Campaign against Headache, HURT questionnaire

Date received: 21 March 2012; revised: 30 October 2012; accepted: 2 November 2012

Introduction

Among adults worldwide, the prevalence of active headache disorder is close to 50%, of migraine 11%, and of tension-type headache (TTH) in excess of 40%. The prevalence of headache on ≥ 15 days/month is 3% (1). About 17% of adults have troublesome headache, causing disability and requiring effective health care, but headache disorders remain 'unrecognized, under-diagnosed and under-treated' (2–4).

Effective health care, meaning individualised and responsive to need, can greatly reduce the personal and societal burdens of headache. Given the cost of these disorders, investment in health care would be

¹Danish Headache Centre, Department of Neurology, Glostrup Hospital, University of Copenhagen, Denmark

²Imperial College London, UK

³Norwegian University of Science and Technology, Norway

⁴Barts and the London School of Medicine and Dentistry, UK

⁵University Consortium for Adaptive Disorders and Head Pain, Italy

⁶Headache Science Centre, Department of Neurology, C. Mondino Foundation, University of Pavia, Italy

⁷Department of Neurology, Albert Einstein College of Medicine, USA

⁸Montefiore Headache Center, Montefiore Medical Center, USA

Corresponding author:

Rigmor H Jensen, Dansk Hovedpine Center, Glostrup Hospital, Nordre Ringvej 67, Glostrup 2600, Denmark.

Email: rigmor.jensen@regionh.dk

sensible (4), but much can be achieved without additional resources by improving efficiency with which resources already committed are put to use. Given the numbers of people with headache, and the fact that, for most people affected, specialist intervention is unnecessary, management of headache belongs mostly in primary care (3–5). It is a fact that, throughout the world, doctors in training receive little teaching on headache (4). Consequently, primary-care physicians are mostly lacking in the knowledge required to manage headache disorders effectively – notwithstanding that diagnosis and management of most people troubled by headache are not difficult.

Lifting The Burden (LTB) is a charitable non-governmental organisation working in official relations with the World Health Organization to conduct the Global Campaign against Headache, initiated in 2003 (6,7). Its over-arching purpose is to reduce the burden of headache disorders worldwide, and it has activities on many fronts and in many countries (2,6,8). The production of management aids, useful at all levels of the health-care delivery system but especially in primary care, and also cross culturally, is one of these fronts (9–11).

A thorough literature search and subsequently a review of 40 ‘psychometrically robust and clinically useful instruments’ showed that there are a number of questionnaires used in headache and/or migraine management (11). These instruments are used for headache diagnosis; assessment of disability, burden and impact; assessment of triggers, exacerbating factors and comorbidities; treatment and follow-up. The Headache Under-Response to Treatment (HURT) Questionnaire (Appendix) (12) is unique as an instrument designed to help non-expert clinicians in primary care improve management of headache. It has been in development since April 2006 by an expert consensus group from all six world regions, with the goal that the instrument must be brief, simple, flexible and useful across cultures and languages. HURT is an outcome measure designed for the one-to-one encounter between health-care provider and patient. It is intended to have utility across the range of common headache disorders and to be informative in two ways for the benefit of the patient: a) by indicating when outcome is less than optimal (in the context of available resources) and b) by suggesting what changes in management might lead to improvement.

Methods

Item development, item reduction and psychometric testing were carried out among 1691 headache sufferers in the United States in 2010. This process assessed criterion validity, as scores on HURT ‘correlated strongly and in the expected direction’ with well-validated

clinical instruments used in assessment of disability (Migraine Disability Assessment Scale, Migraine Prevention Questionnaire), quality of life, and headache impact (HIT-6) (13,14). Pilot testing in the primary-care setting was carried out with a small sample of 40 patients using the Arabic version of HURT (15).

The current version of HURT consists of eight questions to be administered during the course of intervention. The first three questions (HURT-3) relate to frequency of and disability caused by the headache disorder(s) being treated, and the last five (HURT-5) to different aspects of management (medication use and its effects, perception of headache ‘control’, and understanding of diagnosis). HURT might be used at baseline, but this is not its purpose. Responses are graded according to whether they are indicative of change needed in management. At any time during intervention, it should provide the guidance referred to above.

The objectives of this study were to assess a) test-retest reliability of HURT and b) responsiveness to treatment-induced change. Specifically, would HURT show improvement in scores when a headache specialist deemed that best possible outcome (BPO) had been achieved after a treatment period? The hypothesis was that, if it failed on (a) or (b), HURT would not have clinical utility. For this purpose, HURT was intentionally not used as it would be in clinical practice, but applied in specialist care where it could be assumed that treatment of each patient would be optimal.

Project design

For definitions of the terms *reliability*, *responsiveness*, *validity* and *interpretability*, we used the recommendations of the consensus-based standards for the selection of health measurement instruments COSMIN study (16), which represents an international consensus on standardised terminologies used for evaluating health instruments. *Reliability* is defined as ‘the extent to which scores for patients who have not changed are the same for repeated measurement’. An aspect of reliability is *internal consistency*, defined as ‘the degree of interrelatedness among items’. *Responsiveness* is defined as ‘the ability of an instrument to detect change over time in the construct to be measured’.

Consecutive adult patients were recruited from those seeking treatment for a headache disorder at any one of three specialist headache centres in three countries (City of London Migraine Clinic, UK; Danish Headache Centre; Department of Neurology, C. Mondino Foundation, Italy). The only exclusion criteria applied a priori was when an opinion was reached that the patient should be seen urgently, and not after the usual one-month waiting period. Each centre aimed to recruit a minimum of 50 adult patients.

HURT was translated from English into local languages (Danish and Italian) using LTB's translation protocol for hybrid documents (17). Approximately one month before the scheduled first visit, patients on the waiting list to be seen were sent HURT, with an explanation of its purpose and use, and asked to return it by mail in a stamped, addressed envelope. This was designated "pre-visit". Upon arrival at the first scheduled appointment, they were asked to complete it again whilst in the waiting room ("initial visit"). They did not have access at this time to their first questionnaire.

Headache experts thereafter diagnosed and managed the patients according to best practice, European principles of management of common headache disorders in primary care (18), and relevant national treatment guidelines. Patients were followed up for as long as was considered necessary until, in the opinion of the headache experts, the BPO had been achieved. During project design, it was anticipated that BPO would be achieved within one to six months in each patient: quite quickly in, for example, some cases of cluster headache, but longer (>three months) in migraine requiring prophylaxis, in chronic TTH or in medication-overuse (MOH) headache. Upon achieving BPO, patients were asked to complete the questionnaire a third and last time ("final visit") while at the clinic. Only patients who had completed the final questionnaire by 31 December 2010 were included in the study.

Ethics approval

As a service-improvement project, the project fell outside the scope of research ethics review in Denmark and the United Kingdom. Ethics approval was requested in Italy and was granted by the local ethics committee.

Data analysis

Responses to questions 1 to 7 were analysed as ordinal data. Values for question 7 were reversed so that a lower score denoted a better outcome, to be consistent with the other questions. Non-response to question 8 was considered a 'no'.

Intra-rater (test-retest) reliability was assessed using Spearman's correlation coefficient and the Kappa statistic. Linear weights were applied for ordinal data (questions 1 to 7) because we considered not only the difference between two responses, but also the degree by which they were different. For example, the difference between 'always' and 'often' was assumed to be less than the difference between 'always' and 'never'. The weights were applied to reflect greater disagreement (heavier weight, higher impact) for responses that were farther apart (19). Question 8 (yes/no)

produced nominal data so agreement was calculated with unweighted kappa.

Cronbach's alpha was determined for different combinations of questions to gauge internal consistency reliability.

In analysing clinical outcome, the responses were scored according to the four gradations indicating whether change was needed in management; in HURT, these are colour-coded: white (good headache control, no action needed), light grey, medium grey and dark grey (increasingly disabling and inadequately treated headache; action required). Responses in these four gradations were expressed on the scale 0–3, except for questions 6 and 8, which yielded dichotomous responses and were coded 0 or 3. We used this scoring system to give not only appropriate weight to the responses for each question, but also equal contribution of each question to the total score, which has a maximum of 24 (Table 1).

Scores at the initial and final visits were compared using Wilcoxon matched-pairs signed-ranks test for ordinal data, and Chi square test for nominal data.

HURT scores were computed as summations of responses to HURT-3, HURT-5 and all eight questions (HURT-8). Patients who showed a decrease in HURT score were considered clinically improved; those who showed no change or an increase were considered not improved. Wilcoxon's test was used to compare these scores across headache diagnoses. Linear regression analyses were done to see how the scores were predicted by sociodemographic characteristics, duration of headache and diagnosis.

Not included in the HURT questionnaire, but essential to this analysis, were data on the patient's age, gender, education and number of years since onset of headache. These were retrieved from standard patient files. Proportions were used to summarise nominal data. Mean, range and standard deviation were used to summarise continuous variables. Characteristics of patients seen in the three centres were compared using analysis of variance (ANOVA) for age and duration of symptoms, and Kruskal-Wallis ANOVA for educational level.

We accessed the electronic records of the Danish patients to verify dates of clinic visits and intervals between each questionnaire. This also allowed an analysis of the characteristics of dropouts in terms of age, gender and total treatment time.

SPSS 19 and MedCalc 12.3.0 were used to analyse the data.

Results

A total of 291 patients completed the first questionnaire (Denmark 143, UK 103, Italy 45). Pre-visit, initial visit

Table 1. HURT scoring system.

1. On how many days in the last month did you have a headache?	0	0	0	2	3
2. On how many days in the last three months did your headaches make it hard to work, study or carry out household work?	0	1	2	3	3
3. On how many days in the last three months did your headaches spoil or prevent your family, social or leisure activities?	0	1	2	3	3
4. On how many days in the last month did you take medication to relieve a headache? (Do not count preventive medication.)	0	0	1	2	3
5. When you take your headache medication, does one dose get rid of your headache and keep it away?	0	0	1	2	3
6. Do you feel in control of your headaches?	0	0	3	3	3
7. Do you avoid or delay taking your headache medication because you do not like its side effects?	0	0	1	2	3
8. What have you been told is your diagnosis? Do you understand this diagnosis?				0	3

HURT: Headache Under-Response to Treatment Questionnaire; HURT-3 is the sum of scores for the first three questions (min 0, max 9); HURT-5 is the sum of scores for questions 4 to 8 (min 0, max 15); HURT-8 is the sum of all scores (min 0, max 24). The scores correspond to the white-to-grey scale on the questionnaire (see Appendix).

and final visit questionnaires were administered to all patients from Denmark and the UK. Patients from Italy received only the initial visit and final visit questionnaires because of the way in which waiting lists were administered. There were 161 paired pre- and initial visit questionnaires, and 201 paired initial and final visit questionnaires. Because some questionnaires were not completely filled out, paired analysis could not be done for some responses. Participation and dropout are illustrated in Figure 1.

From among the 201 patients who completed the study, the gender ratio was 3:1 (152 females, 49 males). Ages ranged from 17 to 92 years (mean 42.7, SD 14.1 years) and duration of symptoms from 0.1 to 54 years (mean 17.2 years, SD 13.8). There was no significant difference in the duration of illness between patients seen in the three centres ($p=0.15$). Patients from Italy were significantly younger than in other centres ($p=0.01$) while those from the UK reported significantly higher educational levels ($p < 0.001$).

Test-retest reliability was assessed through an analysis of 161 pairs of pre-visit and initial visit questionnaires. Kappa scores ranged from 0.38 to 0.62. Spearman correlation coefficients ranged from 0.49 to 0.76 for questions 1 to 7, with all correlations significant at the 0.001 level (Table 2).

Cronbach's alpha was calculated to measure internal consistency reliability using different combinations of questions. Alpha was calculated for all questions ($\alpha=0.70$ for initial visit and 0.85 for final visit questionnaires), with best results when only questions 1 to 3 were analysed together ($\alpha=0.84$ for initial visit and 0.90 for final visit). Alpha was, as expected, lower for the last five questions ($\alpha=0.30$ for initial visit and 0.68 for final visit).

The time elapsed between the initial and final visits was calculated for the Danish patients. For these 108 patients, the average treatment duration was 13.5 months (SD 6.9 months, range five weeks to 25.6 months; median 13.9 months). The duration of treatment was not related to diagnosis of migraine ($p=0.33$), TTH ($p=0.53$), MOH ($p=0.43$) or post-traumatic headache ($p=0.22$).

Analysis of paired responses to initial and final visit questionnaires of patients from all three centres showed trends towards lower median and mean scores, and significant differences ($p < 0.001$) for all questions except question 7 ($p=0.18$) (Table 3).

Of the patients who responded to question 8 ("Do you feel you understand your diagnosis?") at the final visit, most ($n=141$) identified their diagnosis as migraine; 46 patients indicated TTH and 18 MOH. The percentage of patients who could not write down their diagnosis decreased from 19% to 10% at initial and final clinic visits, whilst the percentage who felt they understood their diagnosis increased from 64% to 87%.

HURT-3, the sum of responses to questions 1–3 (minimum 0, maximum 9), had the highest internal consistency reliability ($\alpha=0.90$); they focus on symptom burden. Comparison of scores at initial and final visits showed a significant change ($p < 0.001$) towards improvement: median decreased from 6 to 4 and mean from 6.03 to 4.36. Range of improvement was -1 to -9 points in the 111 improved patients (55%); in those with worse outcomes (34 patients, 17%), changes ranged from $+1$ to $+9$ points. There was no change in 55 patients (28%).

HURT-5 is the sum of responses to questions 4–8 (minimum 0, maximum 15). There was also a clear difference ($p < 0.001$) and a shift towards lower scores

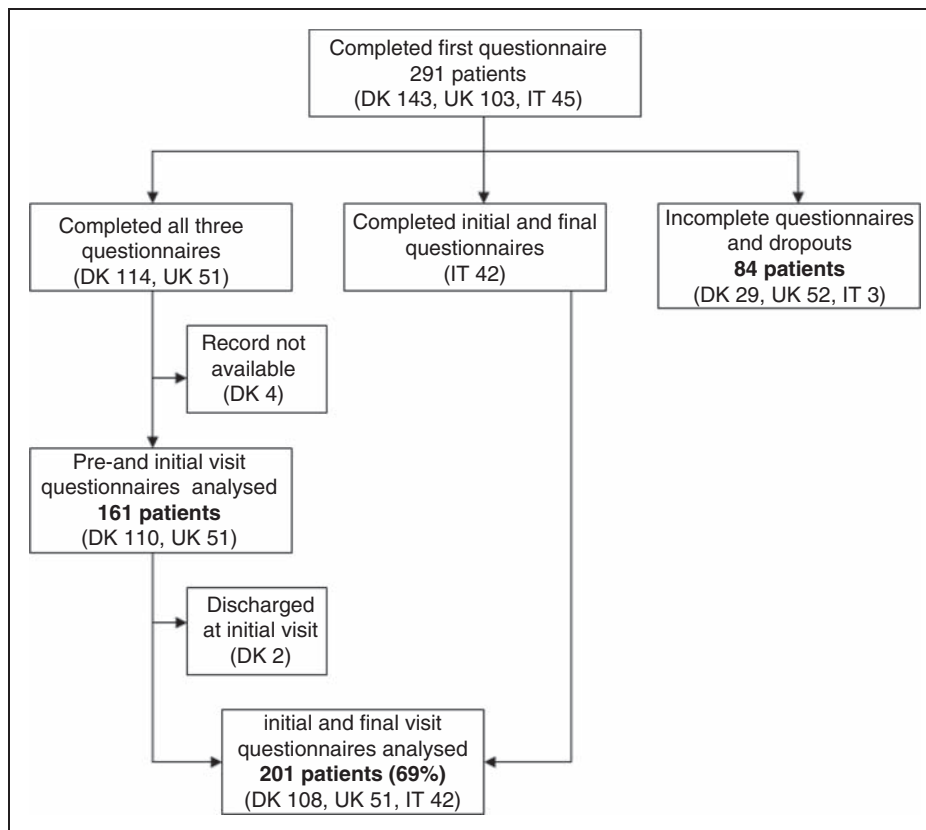


Figure 1. Participation of patients from three headache specialist centres. DK: Denmark; UK: United Kingdom; IT: Italy.

Table 2. Analysis of pre-visit and initial visit responses to HURT questionnaire.

	Number of paired responses	r_s	Kappa (95% CI)
Question 1	161	0.76	$\kappa_{LW} = 0.62$ (0.52–0.72)
Question 2	159	0.51	$\kappa_{LW} = 0.40$ (0.30–0.51)
Question 3	161	0.55	$\kappa_{LW} = 0.44$ (0.34–0.54)
Question 4	152	0.67	$\kappa_{LW} = 0.54$ (0.45–0.64)
Question 5	148	0.49	$\kappa_{LW} = 0.41$ (0.29–0.53)
Question 6	151	0.51	$\kappa_{LW} = 0.39$ (0.28–0.50)
Question 7	141	0.60	$\kappa_{LW} = 0.48$ (0.38–0.59)
Question 8	161		$\kappa = 0.38$ (0.24–0.52)

HURT: Headache Under-Response to Treatment Questionnaire; CI: confidence interval; r_s : Spearman rank-order (rho) correlation coefficient. All correlations significant with p value <0.001; κ : Kappa score for nominal data; κ_{LW} : Kappa score for ordinal data with linear weights.

(decreases in median from 7 to 4 and in mean from 7.05 to 4.07) between initial and final visits. Most patients (75%) showed improvements ranging from -1 to -11 ; 12% had higher final scores (by $+1$ to $+7$ points), while another 13% were unchanged.

Table 3. Medians and means for responses to the HURT questionnaire at initial and final visits.

	Median		Mean	
	Initial visit	Final visit	Initial visit	Final visit
Question 1	2	2	1.87	1.27 ^a
Question 2	3	1	2.17	1.59 ^a
Question 3	2	1	1.99	1.51 ^a
Question 4	1	1	1.32	0.69 ^a
Question 5	2	1	1.50	0.73 ^a
Question 6	3	0	2.28	1.49 ^a
Question 7	1	0	0.94	0.82
Question 8	0	0	1.07	0.40 ^a

HURT: Headache Under-Response to Treatment Questionnaire; ^a $p < 0.001$ for all questions except question 7 ($p = 0.18$). Range is 0 to 3, with lower scores indicating better clinical outcome.

HURT-8 is the sum of responses to questions 1–8 (minimum 0, maximum 24). It also showed a significant difference between paired scores ($p < 0.001$) at initial and final visits: median fell from 14 to 9 and mean from 13.11 to 8.41. The 139 patients (77%) who improved showed decreases ranging from -1 to -19 points,

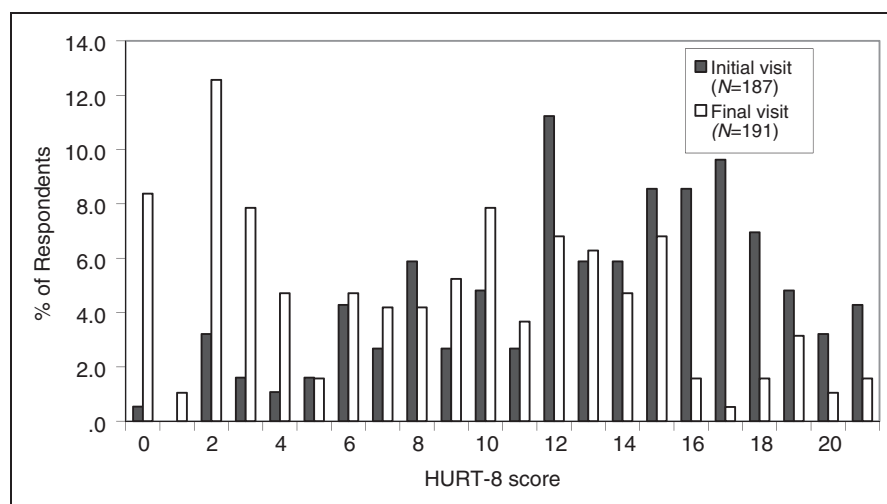


Figure 2. Distribution of patients according to HURT-8.

HURT: Headache Under-Response to Treatment Questionnaire; range of possible responses to HURT-8 is 0 to 24, with lower scores indicating better clinical outcome.

Table 4. Improvement in median (and mean) HURT scores according to diagnosis.

Diagnosis	HURT-3		HURT-5		HURT-8	
	Initial visit	Final visit	Initial visit	Final visit	Initial visit	Final visit
Migraine (141 patients)	6 (5.8)	3 ^c (3.7)	7 (6.6)	4 ^c (3.6)	13 (12.4)	7 ^c (7.3)
Tension-type headache (46 patients)	6.5 (6.0)	5 (4.9)	8 (6.7)	4 ^c (3.5)	13 (12.7)	8 ^c (8.4)
Medication-overuse headache (18 patients)	9 (8.1)	8 ^a (6.7)	10 (8.9)	6 ^b (5.9)	17 (17.0)	13 ^b (12.5)
Chronic post-traumatic headache (7 patients)	9 (8.9)	9 (8.6)	7 (8.1)	6 ^a (5.7)	16 (17.0)	15 ^a (14.3)

HURT: Headache Under-Response to Treatment Questionnaire; ranges are 0 to 9 for HURT-3; 0 to 15 for HURT-5; and 0 to 24 for HURT-8 with lower scores indicating better clinical outcome. ^c $p < 0.001$, ^b $p < 0.01$, ^a $p < 0.05$.

whilst 22 (13%) worsened by +1 to +7 points. The remaining 20 (11%) had unchanged scores. The distribution of patients according to HURT-8 scores is shown in Figure 2.

Table 4 shows how the sum-scores changed from initial to final visit according to diagnosis. Table 5 shows associations between HURT-3 score and migraine.

Logistic regression analysis of HURT-5 and HURT-8 showed that age, gender, treatment centre, education and duration of headache were not significant predictors of clinical outcome, but higher age and shorter duration of headache were significant predictors of HURT-3 ($p < 0.05$). Further analysis of HURT-3 showed that a diagnosis of migraine was a significant predictor of better outcome ($p < 0.05$) when controlled for demographic factors and duration of headache. Regression analyses were not done for other types of headache with few patients (TTH, MOH, post-traumatic headache, trigeminal autonomic cephalalgia, new daily-persistent headache and secondary headaches).

Table 5. Association between migraine diagnosis and clinical improvement indicated by HURT-3 score.

Migraine diagnosis	Clinical improvement indicated by HURT-3		
	Yes	No	N
Yes	89 (64%)	51 (36%)	140
No	12 (29%)	29 (71%)	41
Total	101 (55%)	80 (45%)	181

HURT: Headache Under-Response to Treatment Questionnaire; Chi square test $p = 0.0001$.

Analysis of dropouts and non-responders was done for the Danish patients. The 33 patients with incomplete or unavailable questionnaires were not significantly different from the 110 respondents with complete records in terms of age ($p = 0.28$) or gender distribution ($p = 0.15$). From among the 17 patients who did not complete the last questionnaire, 12 were eventually discharged after the study period.

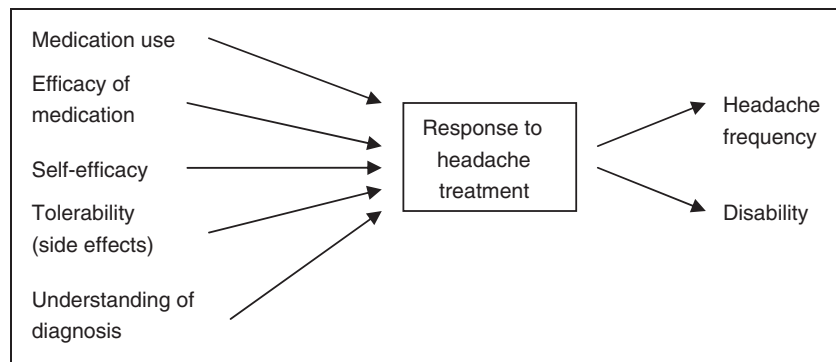


Figure 3. Conceptual framework of HURT showing relationships between variables and construct. HURT: Headache Under-Response to Treatment Questionnaire.

These patients were seen for much longer periods (mean 25 months, SD nine months, $p < 0.001$) compared to the group that completed all questionnaires. The other five patients were lost to follow-up.

Discussion

In evaluating test-retest reliability of the questionnaire, we assumed that the interval between pre-visit and initial visit questionnaires was one month (based on average waiting times at the clinics). Review of the records of the Danish patients showed that the range was larger: from within one day to nine months. There were five patients who filled out the questionnaire on the same day as the clinic visit even though they received the questionnaire much earlier. Most patients were seen at the clinic within two months (median 1.7 months). Longer durations were most likely due to patients' requests for other appointment dates, rather than an unusually long waiting time. We attempted, therefore, a sub-analysis of 37 patients who completed the questionnaires within an interval of one to three months. Their kappa values ranged from 0.26 to 0.78, with only small improvements seen in questions 1 and 8.

In analysing test-retest reliability, we made the assumption that the patients' headache conditions were stable. However, this might not have been true even in the short term. Many of these patients had refractory headaches, and it is possible that their headache characteristics were very unstable, especially during the period before specialist management was started. It is possible that the low test-retest reliability, especially for questions 5, 6, 7 and 8, points toward the dynamic nature of many aspects of headache symptomatology in the short term, and how patients respond to these in terms of medication adjustment and self-education. If it were to be used for baseline assessment, the questionnaire would probably be most helpful to the clinician if it shows the state of the patient immediately prior to an intervention.

To assess the internal consistency reliability of HURT, and to interpret the Cronbach's alpha values, we looked at the questionnaire's conceptual framework (Figure 3). HURT uses a combination of reflective and formative models (20) to describe the construct 'response to headache treatment'.

The right side of the figure (corresponding to HURT-3) is the reflective aspect of the model where a change in the construct is expected to produce change in all the items (effect indicators). On the left side of the figure (corresponding to HURT-5) is the formative aspect of the model where the construct is measured according to five factors. In a formative model, these items (causal indicators) are not interchangeable and cannot replace each other.

The first three questions of HURT (HURT-3) assist the clinician in assessing the frequency of headache and the level of disability caused by it. These three questions have high internal consistency reliability and, as expected, suggest strongly that disability increases with headache frequency.

We must be careful not to conclude that the lower Cronbach's alpha for HURT-5 means that the items should be changed. Rather, it could mean that the five variables – medication use, medication efficacy, tolerability, self-efficacy, and understanding of diagnosis – contribute different parts of the whole construct and do not necessarily correlate with each other. In our study, we found that patients in general continued to be concerned about side effects of medications (question 7 on tolerability) even though there were improvements in the other items.

Interpretability is defined by the COSMIN panel as 'the degree to which one can assign qualitative meaning – that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores' (16). Interpretability of scores is challenging for new instruments going through the iterative process of development and testing. There are two ways to interpret HURT: in terms of single scores and in terms of change in scores over time.

During each clinic visit, clinicians can use the recommendations on the questionnaire to guide the next steps in management. Patients can look at the white-to-grey colour coding to get a general idea of their current status in terms of worst- and best-case scenarios at each point in time.

We saw that, even with best care from headache specialists, patients did not improve in all areas (particularly in their concerns about side effects). There were patients who showed slightly worse scores (13%), or unchanged scores (11%), on HURT-8. This showed us that calculating total scores may be useful at the group level because it allowed us to see how effective treatment moved most of our sample of patients from the higher end to the lower end of the scale (Figure 2). It may also be possible to evaluate response to treatment according to diagnosis: Patients with migraine reported the biggest change (Table 4).

In the clinical setting, however, clinicians and patients should discuss each question separately, because the goals of treatment might be understood differently. Patients might place more importance on some goals over others, and have different ideas of what constitutes effective treatment. For example, they might be willing to tolerate side effects of a new drug (worsening in question 7) if it would mean less headache-related disability (improvement in questions 2 and 3). A slightly increased number of headache days (worsening in question 1) might be acceptable if they felt that they were more in 'control' of their headache (improvement in question 6) or more able to optimise acute medication (question 5).

In this sense, what the questionnaire detects as overall worsening or lack of improvement must be interpreted with care. Rather, in clinical practice, answers to each question in HURT must be analysed individually to determine which aspect of treatment must be adjusted. Each question represents a different dimension of treatment, and, although we give them equal weights in the sum-score, patients may value them differently. We do not recommend a particular difference in sum-scores as a cut-off point for minimal important change. Instead, HURT should be used by patients and clinicians as a starting point to discuss the goals of treatment and to highlight what aspects of management are perceived most important by patients.

Methodological considerations

Patients from Italy and Denmark were referred from general practitioners or specialists. Patients from the UK did not need referral from primary care. There were likely differences between the patients

in this study and those encountered in primary care. We expected patients in specialised headache centres to have initial scores in the higher end of the scale. We also expected patients in primary care to show a wider distribution of scores. The distribution of scores is important for interpretation of statistical analyses, because reliability parameters, including Cronbach's alpha, tend to be higher in heterogeneous populations (20). It is possible that the more homogenous group we sampled in this study (headache patients requiring referral to specialist centres) had less variation in their responses.

Our patients might have been more aware than primary-care patients of their symptoms and treatment plans. They were expected to give more reliable answers to questions 1 to 4, especially if they kept headache diaries. Since they might already have tried a number of treatments before reaching a specialist centre, side effects of medication were, possibly, tolerable as long as there was marked reduction in headache frequency and disability (HURT-3). Perhaps, from their point of view, the goal was reduction in headache disability, even without change in the number of headache days. Patients in primary care may not react in the same way, and may have different judgements on which aspect of treatment carries more weight.

Analysis of the data was limited by missing information, with some patients not completing the entire questionnaire. In a few cases, this was apparently because patients forgot to turn the page to answer several more questions on the back of the paper (which may be a lesson for design and layout). For the most part, however, it appears that patients were not sure about how to answer some questions. Question 8 (understanding of the diagnosis) appeared to be the most difficult to answer, and there may be a contextual reason: Many patients come to specialist centres without a definite diagnosis, waiting for specialist evaluation. As expected, therefore, there was a better response rate to question 8 in the final questionnaire. It should be possible, in a normal clinical setting, for health-care personnel to encourage patients to answer all questions, and to clarify any when needed, such as what it means 'to be in control of your headache' (question 6), which otherwise probably invites very individual and diverse replies. A detailed written instruction regarding problematic questions may or may not add value to the questionnaire: Long written instructions are commonly ignored, or read cursorily.

This study was not designed to look at how physicians acted on the information gained. We cannot estimate the degree to which guidance offered by HURT affected the way headache experts managed their

patients to achieve BPO. The purpose was to observe change in HURT scores, and assess responsiveness to change in clinical conditions, in a group of patients assumed to have been managed optimally. A difficulty here was that, in specialist care, where such a study must be done, there were many patients partially or wholly refractory to best care, and who showed no change or clinical worsening according to HURT scores. Highly significant changes were seen, nonetheless, in seven of eight response-pairs, with a majority experiencing improvement, especially those diagnosed with migraine. Despite this, there appeared to be no change in patients' attitudes towards medications and their side effects. This outcome is very satisfactory with respect to the latent risk of MOH and complexity of patients in specialised headache centres, and is highly encouraging.

Future work

As envisaged, validation in primary care must now follow, with HURT used as intended: to guide management. In the primary-care setting, with patients who may be consulting for the first time, health-care workers can use the questionnaire as an aid in history-taking. Perhaps repeated administration of the questionnaire will have the effect of making primary-care patients more aware of their condition, thereby making them more reliable informants over time. We recommend that HURT be used several times during the course of treatment, not only at baseline and discharge.

The questionnaire may help patients understand that management of their headache proceeds along many fronts, not solely seeking a reduction in headache days. We cannot easily measure how HURT assists patients in reconceptualising their headache management. Perhaps they come to the clinics expecting improvement in terms of symptoms and disability (measured by HURT-3) without having considered the domains in HURT-5. But, after answering the questionnaire, they realise that they must reframe their

expectations of how the management of their headache will proceed.

Translations from English to Danish, Italian and Arabic have been made following LTB's translation protocol. Future work on HURT can focus on assessing cross-cultural validity, defined as 'the degree to which the performance of the items on a translated or culturally adapted patient-reported outcome instrument are an adequate reflection of the performance of items in the original (English) version of the instrument' (16).

In practice, and in line with the concurrent study in Saudi Arabia using the Arabic translation (21), we found HURT easy to use and easy to review.

Conclusion

The HURT Questionnaire is demonstrated to be a useful instrument. It helps patients describe headache frequency and headache-attributed disability, medication use, efficacy and tolerability, and self-efficacy and knowledge about headache. It has very good internal consistency reliability, but test-retest reliability was not very high, probably because of unstable headache before the start of specialist management. In seven questions, HURT is responsive to clinical change. HURT may be able to guide physicians in how to proceed with management towards the best possible outcome, and evaluate the quality of management. There was no improvement (at group level) in patients' concerns about side effects of medication, which are to be noted by the physician but may or may not call for change in management. To evaluate change in symptom burden, HURT-3 can be used; to evaluate headache management, including medication, self-efficacy and knowledge about headache, HURT-5 is appropriate. HURT-8 gauges clinical outcome overall. We recommend that a computerised version be developed for future daily practice. Further evaluation and clinical testing of HURT in primary-care settings are needed.

Clinical implications

- The Headache Under-Response to Treatment (HURT) Questionnaire is a unique instrument designed to help non-expert clinicians in primary care improve management of headache by measuring effectiveness of intervention and giving suggestions towards improvement.
- Earlier work has been done on the questionnaire's criterion validity, with promising results.
- This original research paper is a first assessment of clinical utility conducted in headache-specialist centres in three countries in order to demonstrate that HURT was responsive to change induced by effective management, and to assess test-retest reliability.

Funding

This study is part of the Global Campaign against Headache and supported by Lifting The Burden. The project received no grant from any funding agency in the commercial sector.

Conflict of interest

Timothy Steiner and Rigmor Jensen are directors of Lifting The Burden. The other authors declare no conflict of interest that is of relevance to this work.

References

1. Stovner LJ, Hagen K, Jensen R, et al. The global burden of headache: A documentation of headache prevalence and disability worldwide. *Cephalalgia* 2007; 27: 193–210.
2. World Health Organization. *Headache disorders. Fact sheet Number 277*. Geneva: WHO, <http://www.who.int/mediacentre/factsheets/fs277/en/> (2004, accessed 6 March 2011)
3. World Health Organization. *The World Health Report 2001*, Geneva, <http://www.who.int/whr/2001/en/index.html> (2001, accessed 7 March 2011).
4. World Health Organization. *Atlas of headache disorders and resources in the world 2011*, http://www.who.int/mental_health/management/who_atlas_headache_disorders.pdf (2011, accessed 8 June 2011).
5. World Health Organization Department of Mental Health and Substance Dependence. *Headache disorders and public health: Education and management implications*. Geneva: WHO, 2000.
6. Steiner TJ. Lifting The burden: The global campaign against headache. *Lancet Neurol* 2004; 3: 204–205.
7. Lifting The Burden: The global campaign to reduce the burden of headache worldwide, <http://www.l-t-b.org> (accessed 14 August 2012).
8. Steiner TJ, Birbeck GL, Jensen R, et al. Lifting The Burden: The first 7 years. *J Headache Pain* 2010; 11: 451–455.
9. Steiner TJ and Martelletti P. Aids for management of common headache disorders in primary care. *J Headache Pain* 2007; 8: S2.
10. Steiner TJ. The HALT and HART indices. *J Headache Pain* 2007; 8(Suppl 1): S22–S25.
11. Buse DC, Sollars CM, Steiner TJ, et al. Why HURT? A review of clinical instruments for headache management. *Curr Pain Headache Rep* 2012; 16: 237–254.
12. Lifting The Burden. HURT Index Questionnaire (v 2.2), http://www.l-t-b.org/assets/96/AB9C64EE-D800-01AB-1FAC446E25B0DF88_document/HURT_NY_revision.pdf (undated, accessed 9 August 2012).
13. Buse DC, Steiner TJ, Serrano D, et al. The Headache Under-Response to Treatment (HURT) index: An outcome measure to improve management of headache, especially in primary care. *Neurology* 2010; 74(Suppl 2): A323.
14. Buse DC, Steiner TJ, Serrano D, et al. Psychometric validity of the lifting the burden Headache Under-Response to Treatment (HURT) Index: A new instrument developed to improve management of headache. *Headache* 2010; 50: S42–S43.
15. Al Jumah M, Al Khathaami A, Kojan S, et al. HURT Index in primary care: Reliability and impact on headache management and patients' satisfaction. *J Neurol* 2010; 257: S98.
16. Mokkink LBW, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63: 737–745.
17. Lifting The Burden. *Translation protocol for hybrid documents*, http://www.l-t-b.org/assets/0/AD0EF573-E65D-8DD4-B7F22842405C1D95_document/Translation_protocol_-_Hybrid.pdf (accessed 27 October 2012).
18. Steiner TJ, Paemeleire K, Jensen R, et al. European principles of management of common headache disorders in primary care. *J Headache Pain* 2007; 8: S3–S47.
19. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bull* 1968; 70: 213–220.
20. de Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in medicine: A practical guide*. New York: Cambridge University Press, 2011.
21. Al Jumah M, Al Khathaami A, Tamim A, et al. Reliability, validity and clinical utility of the Arabic version of the HURT (Headache Under-Response to Treatment) Index in the management of primary headache disorders. Report, 2011.

Appendix

Lifting The Burden

in official relations with the World Health Organization

The Global Campaign against Headache

HURT Questionnaire

(Headache Under-Response to Treatment)

Your medical treatment for your headaches may not be as good as it can be. By completing this short questionnaire, you will help your doctor or nurse improve it.

Please answer these questions carefully

please tick **ONE** box in each row

1	On how many days in the last month did you have a headache?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		none	1-2	3-5	6-15	16+
2	On how many days in the last three months did your headaches make it hard to work, study or carry out household work?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		none	1-5	6-10	11-20	21+
3	On how many days in the last three months did your headaches spoil or prevent your family, social or leisure activities?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		none	1-5	6-10	11-20	21+

Analysis (these questions establish frequency of all headaches and of disabling headaches under current treatment; ticks towards the right suggest increasing need for treatment review)

All ticks in white area	Headache control is good: no review needed.
One or more ticks in lightly-shaded area	Better acute headache management is needed; review Qs 4-8 for guidance; prophylaxis may not be required.
One or more ticks in middle-shaded area	Headache control is not good; review Qs 4-8 to optimise acute medication; consider ways of reducing frequency (trigger avoidance and prophylactic medication).
One or more ticks in dark-shaded area	Disabling headache, poorly treated; possibly chronic daily headache (acute medication should be avoided); review Qs 4-8 and consider ways of reducing frequency.

4	On how many days in the last month did you take medication to relieve a headache? (Do not count preventative medication.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		none	1-4	5-9	10-15	16+	
5	When you take your headache medication, does one dose get rid of your headache and keep it away?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		always	often	sometimes	rarely	never	
6	Do you feel in control of your headaches?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		always	often	sometimes	rarely	never	
7	Do you avoid or delay taking your headache medication because you do not like its side-effects?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		never	rarely	sometimes	often	always	
8	What have you been told is your headache diagnosis?	please write your diagnosis here:				<input type="checkbox"/>	<input type="checkbox"/>
	Do you feel you understand this diagnosis? [tick one box]					yes	no

Analysis (these questions suggest how current management might be improved)

<p>Q4: Response should accord with Q1. When medication days are 5-9 there is potential risk of medication overuse. When medication days are >10 there is high risk of medication-overuse headache.</p> <p>Q5: Ticks towards the right increasingly suggest poor efficacy</p> <p>Q6: This question relates to self-efficacy and to satisfaction.</p> <p>The response should be concordant with previous responses.</p> <p>Q7: Ticks towards the right increasingly suggest poor tolerability.</p> <p>Q8: This question relates to education.</p>	<p>Advise patient about the risk and dangers of medication overuse. Give written information leaflet.</p> <p>Consider ways to reduce frequency (trigger avoidance and prophylactic medication).</p> <p>Consider treating earlier, changing medication, dose or route of administration, or using combination therapy, according to local guidelines.</p> <p>When the response is in the shaded area, look for the reason(s) in responses to Qs 1-6. If it is not evident, consider the possibility of co-morbidities.</p> <p>When the response is not concordant, consider cognitive interventions and expectation management.</p> <p>Consider changing medication or dose according to local guidelines.</p> <p>Always hand out the appropriate information leaflet. When the diagnosis is wrongly stated, or the answer "no" is given, further explanation may be necessary.</p>
---	--

The Headache Under-Response to Treatment (HURT) Questionnaire. This can be downloaded from Lifting The Burden's website: www.l-t-b.org © Lifting The Burden.