CHESS PROMs review – Cephalalgia format

Manuscript Title:

Assessing the impact of headaches and the outcomes of treatment: a systematic review of patient-reported outcome measures (PROMs)

Authors:

Kirstie L Haywood[1,4], Tom S Mars[2,4], Rachel Potter[2,4], Shilpa Patel[2,4], Manjit Matharu[3,4], Martin Underwood[2,4], On behalf of the CHESS team[4]

Affiliation:

1. Warwick Research in Nursing, Department of Health Sciences, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.
2. Warwick Clinical Trials Unit, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.
3. Headache Group, UCL Institute of Neurology, Queen Square, London, UK. WC1N 3BG
4. On behalf of the CHESS team. Warwick Clinical Trials Unit, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.

Corresponding author:

Dr Kirstie L Haywood. Senior Research Fellow. Royal College of Nursing Research Institute, Department of Health Sciences, Warwick Medical School, The University of Warwick, Gibbet Hill, Coventry. CV4 7AL.

Email: k.l.haywood@warwick.ac.uk

Telephone: 024761 59616

Abstract 198/200; Main text 4099/4000.

Refs 85; Tables 1; Figures 1; Appendices 7.

Structured Abstract: (198/200)

Aims

To critically appraise, compare and synthesise the quality and acceptability of multi-item PROMs for adults with chronic or episodic headache.

Methods

Systematic literature searches of major databases (1980-2016) to identify published evidence of PROM measurement and practical properties. Data on study quality (COSMIN), measurement and practical properties per measure was extracted and assessed against accepted standards to inform an evidence synthesis.

Results

From 10,903 reviewed abstracts, 103 articles were assessed in full; 46 provided evidence for 23 PROMs: eleven specific to the health-related impact of migraine (n=5) or headache (n=6); six assessed migraine-specific treatment response/satisfaction; six were generic measures.

Evidence for measurement validity and score interpretation was strongest for two measures of impact - Migraine-Specific Quality of Life Questionnaire (MSQ v2.1) and Headache Impact Test 6-item (HIT-6), and one of treatment response - the Patient Perception of Migraine Questionnaire (PPMQ-R). Evidence of reliability was limited, but acceptable for the HIT-6. Responsiveness was rarely evaluated. Evidence for the remaining measures was limited. Patient involvement was limited and poorly reported.

Conclusion

Whilst evidence is limited, three measures have acceptable evidence of reliability and validity - HIT-6, MSQ v2.1 and PPMQ-R. Only the HIT-6 has acceptable evidence supporting its completion by all 'headache' populations.

Key words:

headache; patient-reported outcome; validity; reliability; systematic review

Background: ***(313)***

Headache disorders are common in the adult population; the most common - tension-type and migraine - have a one-year prevalence of 40% and 11% respectively [1,2,3]. Between 2-4% of the general population experience chronic headache [4,5]. Headache disorders can profoundly impact an individual's functional ability and quality of life [3,6]. Affecting primarily young adults, the personal and economic burden of headache is substantial and comparable to other chronic conditions such as congestive heart failure, hypertension, or diabetes [7].

An individual's self-report of the presence, severity, frequency and impact of headache is crucial to understanding the effectiveness of therapeutic interventions. Patient-reported outcome measures (PROMs), which seek to provide a patient-based assessment of the impact of headache on how people feel, function and live their lives are now available. Where recommendations to include PROMs in headache clinical trials are available [8,9], specific guidance for PROM-based outcome reporting does not exist. The integrity of PROM-based reporting is underpinned by clear evidence of essential measurement and practical properties in the clinical population of interest [10,11]. It cannot be assumed that the reliability and validity of measure is consistent across different types of headache, and evidence of PROM performance across different sub-types is often not available [12]. PROM score interpretation also requires guidance for what change in score reflects a meaningful change in 'headache' for the individual patient (minimal important change (MIC)) and what difference reflects a meaningful difference between groups of patients defined by some external anchor (minimal important difference (MID)) [10,11]. Structured reviews of PROM performance provide essential evidence to inform the selection of robust, relevant and acceptable measures.

In this systematic review, we critically appraise, compare and synthesise published evidence of essential measurement and practical properties for clearly defined PROMs evaluated in adult headache populations. The review provides a transparent summary of the evidence-base with which to inform PROM selection for future application in headache-specific research.

Methods: (693)

Identification of studies and PROMs: search strategy

The search strategy was developed by experienced reviewers (KH, TM, RP, SP) and with expert librarian support to retrieve references relating to the development and/or evaluation of multi-item PROMs used in the assessment of adults (aged 18 years and above) with chronic or episodic headache including migraine.

Medical subject headings (MeSH terms) and free text searching were used to reflect three characteristics: 1) population – headache and migraine; 2) type of assessment – patient-reported outcome measures (PROMs); and 3) measurement and practical properties [11,13,14]. The full search strategy is available in Appendix 1.1.

Two databases were searched (MEDLINE (OVID), EMBASE (OVID); 1980 to December 2016) (figure 1). A subsequent search incorporated the names of more than 50 multi- and single-item measures identified during the initial search (Appendix 1.2 and 1.3). From a total of 39 multi-item PROMs thus identified, 16 had been superseded by revised measures or were no-longer in use as evidenced by their lack of inclusion in studies published post 2000 (Appendix 2). Given that such measures are unlikely to be of interest, the eligibility criteria for the review and analysis was revised to focus on PROMs 'in use' post-2000.

The citation lists of included articles and existing reviews were also reviewed [15,16]. Named author searches were conducted. *Inclusion/exclusion criteria*

Titles and abstracts of all articles were independently assessed for inclusion/exclusion by two reviewers (TM, KH) and agreement checked. Published articles were included if they provided evidence of development/evaluation for clearly defined, reproducible, multi-item PROMs, following self-completion by adults who self-reported or had been diagnosed by a clinician as having a headache disorder. Articles relating solely to the application of measures without some evidence of measurement and/or practical properties were excluded. Articles describing the translation of PROMs and/or evaluations in non-English speaking populations were also excluded. Conference papers and abstracts were excluded.

Included PROMs must be 'in-use' in research published between 2000-2016. PROMs were categorised as: generic (profile; utility) or condition-specific (headache; migraine). Clinician-reported, diagnostic and screening measures were excluded. Domain-specific measures that were not specific to the impact of headache, and measures that were not clearly reproducible were excluded.

Data extraction and appraisal

A data extraction form was informed by guidance for PROM evaluation [10,11,17], published PROM reviews [14,18,19] and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist [20,21]. The form captured both study and PROM-specific information. Population diagnosis and diagnostic criteria (if any) were extracted. We sought evidence on: reliability (internal consistency; test–retest, intra/inter-tester); validity (content; construct; known groups); responsiveness; interpretation (minimal important change (MIC) and/or difference (MID)); and precision (data quality; end effects). Evidence for the practical properties included acceptability (relevance; respondent burden) and feasibility. Evidence of active patient involvement in PROM evaluation was also sought [18,22,23]. All publications were double-assessed (KH,TM) and agreement checked.

Assessment of study methodological quality

One experienced reviewer (KH) applied the COSMIN checklist to assess the methodological quality of included studies [20,21]. Methodological quality was evaluated per measurement property on a 4-point rating scale (excellent, good, fair, poor) and determined by the lowest rating of any items in each checklist section [21].

Assessment of PROM quality

A similar checklist for PROM quality does not exist. Therefore, a pragmatic checklist informed by a synthesis of various recommendations was adopted [18,19,21,24](Appendix 3: Table 2). To provide a global overview of the concepts captured within the reviewed headache-specific measures, items were categorized per domains of one of the most frequently used conceptual models of health-related quality of life (HRQOL) – the Ferrans revision to the Wilson and Cleary model [25,26].

Data synthesis

A qualitative synthesis of evidence per reviewed PROM per reported measurement property informed the overall judgement of quality and acceptability. The synthesis combined four factors: 1) study methodological quality (COSMIN scores); 2) number of studies reporting evidence per PROM; 3) results per measurement property (Appendix 3: Table 2); and 4) evidence of consistency between evaluations [23,27]. Two elements of the data synthesis are described: First, the overall quality of a measurement property was reported as: adequate (+), conflicting (+/-), inadequate (-), or indeterminate (?). Second, evidence for the overall quality of evidence was categorized: ‘strong’, ‘moderate’, ‘limited’, ‘conflicting’, or ‘unknown’ [27].

Results: (1874)

Identification of studies and PROMs

Study and PROM identification is summarized per PRISMA guidance in figure 1 (www.prisma-statement.org). Forty-six articles provided evaluative evidence for 23 PROMs (Appendices 4 and 5 (Tables 3 and 4)). Six assessed impact of headaches overall the EUROLIGHT [28]; Headache Activities of Daily Living Index (HADLI) [29]; Headache-specific Disability Questionnaire (HDQ) [30]; the Headache Impact Test (HIT) [3] and its short-form HIT-6 [31]; and a headache-specific modification of the Short-Form 36-item Health Survey [32]. Five were specific to the impact of migraine: Functional Assessment in Migraine questionnaire (FAIM) [33]; Headache Needs Assessment Survey (HANA) [34]; MIgraine Disability ASessment (MIDAS) [35]; Migraine-Specific Quality of Life Questionnaire (MSQ v2.1) [36]; and the Migraine-Specific Quality of Life (MSQOL) measure [37]. Six assessed response to and/or satisfaction with migraine-specific drug treatment: Completeness of Response to migraine therapy (CORS)[38]; Migraine Assessment of Current Therapy (Migraine-ACT) [39]; Migraine-Treatment Assessment Questionnaire (M-TAQ) [40]; Migraine-Treatment Optimisation Questionnaire (M-TOQ) [41]; Migraine Treatment Satisfaction Measure (MTSM)[42]; and the Patient Perception of Migraine Questionnaire – Revised (PPMQ-R) [43]. Item content of all specific measures is illustrated in Appendix 6 (Table 5).

Finally, six generic measures had been assessed in headache populations: the Short-Form 36-item Health Survey (SF-36)[44], SF-12 [45], SF-8 [46], EuroQoL EQ-5D 3L [47], Health Utility Index-3 (HUI-3)[48] and the Quality of Well-being Scale (QWB)[49,50].

Patient and study characteristics (Appendix 5 (Table 4))

Patient populations ranged 18 to 83 years, were largely white, often with large proportions of female participants. Sample sizes ranged 25 to more than 8,500. Populations included mixed, chronic and/or episodic headache or migraine. Where clinician-based diagnosis was described, most adopted the International Classification of Headache Disorders (ICHD-II). However, for many, patients were self-diagnosed, and a wide range of diagnostic criteria were described. Most studies were cross-sectional or longitudinal surveys. Nine were clinical trials or involving data secondary analysis. Fourteen studies were specific to PROM development and/or initial evaluations. Most evaluations were with US populations.

Measurement properties and methodological quality

Study methodological quality per measurement property per reviewed PROM is presented in Appendix 7 (Table 6). The overall evidence synthesis is presented in Table 1.

i. PROMs assessing Migraine and Headache-specific impact (n=11)

Apart from the FAIM, MSQ v2.1, MSQoL and HIT, all measures lack a clear description of aim, the concepts being measured or the process of items generation. The FAIM [33], MSQ v2.1 [36] and MSQoL [37] involved expert clinicians and patients in item generation, supporting a positive rating of content validity.

The HIT 'item bank' was informed by four legacy measures – the MIDAS, MSQ (v1.0), Headache Disability Index (HDI) and Headache Impact Questionnaire (HIMQ) – and consultation with clinicians [3]. Apart from the MSQ, item generation for these measures is poorly reported but largely driven by clinical opinion. Additional evaluations of the content validity of the item bank or short form measures is not described. Clinical opinion, literature review, and/or the completion of established questionnaires were the main sources of items for the remaining measures. There was no evidence of active patient collaboration in PROM development and/or evaluation.

The shortest measures are the MIDAS (5-items) and HIT-6 (6 items); the longest is the 103-item EUROLIGHT (Table 2). Apart from the FAIM, all assess headache/migraine symptomology. While five headache-specific measures assess pain, the migraine-specific measures do not. Only the HANA, MSQv2.1 and HIT-6 assess fatigue.

All assess the impact of headache/migraine on social function, activities of daily living and/or work. Seven – FAIM, HANA, MSQv2.1, MSQOL, HIT, HIT-6, EUROLIGHT – assess the emotional burden of headache/migraine; five of these – FAIM, MSQv2.1, HIT, HIT-6, EUROLIGHT – plus the HADLI, assess the impact on cognition and difficulty thinking.

Acceptable evidence of measurement dimensionality from studies of at least moderate methodological quality was reviewed for five measures – FAIM [33], MSQv2.1 [12,51], MSQoL [52], HIT [3], HIT-6 [53]; three have moderate to strong evidence of both structural validity *and* internal consistency – FAIM [33], MSQ v2.1 [12,36,51,54] and the HIT-6 [31,41,53,55,56] (Table 1; Appendix 7). Three measures have acceptable evidence of internal consistency reliability from studies of at least moderate methodological quality, supporting application in the assessment of groups (FAIM)[33] and individuals (MSQ v2.1 [12,36,51], HIT-6 [53,56]) (Table 1; Appendix 7); but for the majority evidence was limited (n=3), from poor quality studies (n=3) or not available (n=1). Only the HIT [31,57] and HIT-6 [31,53,56,57] have acceptable evidence of temporal stability supporting application in the assessment of groups and individuals. Evidence for the remaining measures was limited.

Five measures have acceptable evidence from good quality studies describing their construct validity – FAIM [33], MIDAS [56], MSQ v2.1 [12,36,43,53], HIT [57] and HIT-6 [12,53,56,57]. For the remaining measures evidence was of poor quality (n=4) or not available (n=2); authors often failed to hypothesise a priori the association between variables.

Evidence of responsiveness was limited. Statistically significant between-group differences for average HIT-6 and total HIT change scores were reported for patients categorised per self-reported change (better / same / worse) in physical activity, level of frustration or daily activities following a 3-month follow-up period of 'usual care' [31].

Large and moderate effect size statistics were reported for the MSQv2.1 [12] and HIT-6 [53] in patients who reported large or moderate improvement in the number of headache days following a pharmaceutical-based clinical trial, respectively. Following a non-comparative, observational study of zolmitriptan for an acute migraine attack, small and moderate ES statistics were reported for the SF-36 and MSQoL respectively [52].

Following completion of the HIT-6 by patients with chronic daily headache in a trial of usual medical care (UMC) versus UMC plus acupuncture, an anchor-based estimate of the MIC was calculated as approximately 3.7; the MID was estimated as 2.3 [58]. Change in HIT-6 scores that exceeded the proposed MIC were reported in patients with chronic migraine receiving onabotulinumtoxinA in a placebo-controlled double blind trial; a between group difference that exceeded the MID, in favour of the active treatment, was also reported [59].

Both anchor-based [60,61] and distribution-based estimates [60] were calculated for the MSQv2.1 following completion by patients with chronic migraine. Cole et al [60] proposed an MIC of 5.0 for the RR domain, with ranges for the RP (5.0 to 7.9) and EF (range 8.0 to 10.6) domains; MIDs were recommended as: RR 3.2, RP 4.6, EF 7.5 [60]. A between group difference that exceeded the proposed MID, in favour of the active treatment, was reported for the MSQv2.1 RR domain only in patients with chronic migraine receiving onabotulinumtoxinA in a placebo-controlled double blind trial [59]. However, within-individual change scores were larger than the proposed MIC for each domain for patients receiving active treatment.

ii. PROMs assessing response to or satisfaction with migraine-specific treatment (n= 6 measures)

Four of the six measures - the CORS, M-TOQ, MTSM and PPMQ-R - have acceptable descriptions of the measurement aim, conceptual underpinning and item generation. Although detail is limited,

three measures – CORS, MTSM and PPMQ-R - involved both expert clinicians and patients in item generation (the MTSM involved US and UK participants), supporting a positive rating of content validity; the M-TAQ utilised patient interviews and focus-groups, with additional reference to established treatment-optimisation measures.

Item generation for the M-ACT [39] and the M-TOQ [41] was informed by clinical evidence and the consensus of clinical headache experts and researchers; patients were not involved, supporting a negative rating of content validity. There was no evidence of active patient collaboration.

The shortest measures are the M-ACT (4-items) and M-TOQ-5 (5-items); the longest is the 45-item MTSM (Appendix 4). Apart from the M-ACT and M-TAQ, all assess migraine symptomology, including pain severity, and the wider impact on activities of daily living and/or work; the PPMQ-R also assesses limitations in social functions (Appendix 6). The CORS, M-TOQ-15 and PPMQ-R assess the emotional burden of migraine; just the CORS and PPMQ-R also assess cognition and difficulty thinking. Three measures assess if the patient has 'returned to normal' - CORS, M-ACT, M-TOQ. All assess confidence in/or satisfaction with treatment; the M-TOQ assesses treatment side-effects.

Only the PPMQ-R has acceptable evidence of measurement dimensionality *and* internal consistency reliability from studies of at least moderate methodological quality (Table 1; Appendix 7). For three measures – CORS, M-TOQ, MTSM - evidence was acceptable but limited.

Only the M-ACT has acceptable evidence of temporal stability from several studies of fair methodological quality, supporting application in the assessment of groups (Table 1; Appendix 7). Evidence for three measures - M-TAQ, M-TOQ, PPMQ-R - was limited to single studies judged to be of fair quality (Table 1; Appendix 7). Only the PPMQ-R and MTSM have acceptable evidence of construct validity from good quality studies. For the remaining measures evidence was limited (CORS, M-TAQ, M-TOQ) or from poor quality (M-ACT) studies.

Following a 2-month pharmaceutical trial, small to moderate change score correlations between the CORS and the PPMQ-R supported a priori hypothesised associations, providing acceptable, but limited, evidence of responsiveness [38]. Further criterion-based evidence, comparing the comparative CORS with change in CORS sub-sets at 2-months, provided additional, hypothesis driven evidence of responsiveness [38]. Small to moderate effect size statistics were reported for the PPMQ-R in patients categorised per self-reported improvement (range 0.14 to 0.50) or worsening (range 0.06 to 0.23) in pain severity; the largest ES were reported for the Efficacy and Function domains [43].

The Standard Error of Measurement (SEM) was calculated for the PPMQ-R, as a reflection of the within individual minimal change in score (MIC) [43]. Apart from the Cost domain (SEM 11.0), SEM estimates ranged 3.4 (Bothersome) to 5.4 (Total score), supporting an MIC recommendation of 5 points for the total score and Efficacy, Function and East of Use domains. Results suggest that the Cost domain is highly variable and not responsive to change in migraine severity or role limitation.

Estimates of the minimally important change and minimally important difference were reviewed for three headache-specific measures - MSQ v2.1 [36], HIT-6 [31], PPMQ-R [43]. Completion of the HIT-6 by Dutch patients with chronic tension-type headache [62] and episodic migraine [63] suggested a wider range of MIC values –from -2.5 [63] to -8.0 [62] than that determined in a US population with chronic daily headache (-3.7) [58]. The differences were largely explained by use of different anchors – where a greater perceived change was the imposed anchor, a larger MIC was calculated. An MIC of >8.0 suggests that improvement must be present in at least two of the six HIT-6 items [62], which may be judged a relevant treatment effect [62,63]. Similarly, suggested MID values range from -1.5 (episodic migraine) [63] to -2.3 (chronic daily headache) [58].

iii. Generic PROMs (n= 6)

Evaluations of all generic measures in the headache population were very limited. There was no evidence exploring the content validity or relevance of the six reviewed generic measures with the headache population. There was no evidence of active patient collaboration.

Where applicable, there was no evidence of measurement dimensionality or internal consistency reliability (Table 1). Just one measure – the QWB-SA - had conflicting evidence of temporal stability from one study, judged to be of poor methodological quality [64] (Table 1; Appendix 7).

Acceptable evidence of construct validity from several studies judged to be of fair or good methodological quality, was reviewed for both the SF-36 [36,55,65] and the SF-8 [7,31,57,56]; for the SF-12 evidence was limited (Table 1; Appendix 7). For the remaining measures evidence was limited (EQ-5D) or of poor quality (HUI-3, QWB). There was no evidence of measurement responsiveness.

Discussion (1219)

High quality, relevant and acceptable PROMs provide patient-derived evidence of the impact of headache and the relative benefit of associated healthcare at both the time of the headache and the intervening period. The importance of capturing the patient perspective is reflected in the large number of measures included in this review. However, apart from two condition-specific – HIT-6 and MSQv2.1 - and one treatment-response – PPMQ-R – measures for which strong evidence was reviewed, evidence was largely limited or not available.

This is the first systematic review to include a methodological assessment of both study and PROM quality in the headache population. Clarity in PROM focus is an essential, but often over-looked aspect of PROM development [24,80]. Except for four condition-specific - MSQ v2.1, MSQoL, HIT, HIT-6 - and four treatment-response measures - CORS, M-TOQ, MTSM, PPMQ-R - all lacked a clear description of the measurement aim. Moreover, the condition-attribution of measures was not always self-evident: just three 'migraine-specific' measures assessed the impact of 'migraine' – FAIM, MSQ v2.1, MSQoL. The HANA includes both 'migraine' and 'headache' in the item stem and, despite the name, the MIDAS assesses the impact of 'headache'. It is suggested that the attribution of 'headache' supports a 'broader' assessment than would be achieved with 'migraine'; moreover, many patients may be unaware of a migraine diagnosis [3]. The HIT item content was informed by both migraine (MSQ and MIDAS) and headache-specific (HIMQ, HDI) measures; a content comparison failed to reveal any systematic differences in concept coverage and further evaluation in a mixed population supported the uni-dimensionality of headache disability [3]. Evidence further supports the ability of the HIT to assess headache disability across a wide spectrum of impact, avoiding the potential for ceiling effects, following completion by headache and migraine populations [3,63]. Just four measures - the HIT-6, HADLI, HDQ, MIDAS – have been evaluated in both headache and migraine populations. However, whilst evidence is strong for the HIT-6, the remaining measures should be applied with caution.

Except for two condition-specific – MSQv2.1, MSQoL - and four treatment-response measures – CORS, M-TAQ, MTSM, PPMQ-R - the extent of patient participation was limited and poorly detailed. Moreover, except for three measures – MSQoL, PPMQ-R, EUROLIGHT - PROM relevance, content and face validity was not explicitly explored with patients and/or expert panels. Item content for the remaining measures was informed by a mix of qualitative research with clinicians, reference to existing measures, published literature and/or completed questionnaires. Successful treatment for headache disorders should seek to improve both overall quality of life, as well as an individual's quality of life during the attack [37]; and assessment should seek to capture these distinctions.

Although varying in length, there was a similarity of item content across condition-specific measures. Most assessed headache/migraine-related symptomology; pain severity was commonly assessed by headache-specific and treatment-response measures, but not by the migraine-specific measures. Just two measures - MSQv2.1, HIT-6 – assessed fatigue. Measures with a primary focus on symptomology have been criticised for failing to take into consideration the longer-term consequence of, or fear associated with, a potentially severe headache or migraine, such as evading commitments or making plans [81,82]. Nevertheless, except for the FAIM and HANA, all condition-specific and most treatment-response measures also assessed the wider impact of headache on social function and interactions, activities of daily living and/or work. Several measures - MSQv2.1, HIT, HIT-6, EUROLIGHT, CORS, PPMQ-R - also assessed both the emotional burden and cognitive impact of headache/migraine.

Three condition-specific - FAIM, MSQv2.1, HIT-6 - and one treatment-response - PPMQ-R - measures have strong evidence of both structural validity and internal consistency reliability. Factor analysis supported the uni-dimensionality of the FAIM following completion by migraineurs, and the HIT-6 as a measure of 'headache disability' following completion by mixed populations. A three-domain structure of the MSQv2.1 was supported – Role Restriction (RR), Role Prevention (RP) and Emotional Function (EF) - following completion in both chronic and episodic migraine populations. However, for most measures evidence of structural validity or internal consistency reliability was limited, from methodologically poor quality studies or not available. Evidence of temporal stability was also limited, and available only for the HIT, HIT-6, M-ACT, M-TAQ, M-TOQ and PPMQ-R. There was no evaluation of measurement error.

Five condition-specific (FAIM, MIDAS, MSQ v2.1, HIT, HIT-6), two treatment-response (MTSM, PPMQ-R) and two generic (SF-36, SF-8) measures have acceptable evidence of construct validity from good quality studies. For the remaining measures, evidence was limited, of poor methodological quality or not available. Methodological inadequacies included small sample sizes and a failure to hypothesise a priori the expected association between variables. As reported in other reviews [18,19], there was limited evidence of responsiveness: just two studies [31,38] provided acceptable, but limited, evidence for the CORS and HIT measures. Evaluative measures require evidence of responsiveness to demonstrate that they can detect real change in condition over time; without such evidence measures should be applied with caution.

Whilst a limitation of the review is that we have only included evaluations in English, the context, setting and population are important in appraising evidence of PROM measurement and practical properties [83]. Moreover, the diversity of reviewed measures reflects the wide range of assessment

approaches in current use. Reviewed studies were of adults aged 18 years and over; with no upper age-limit imposed. All reviewed studies excluded people with significant co-morbidities. We are confident that the results are generalizable to the wider population of English-speaking adults with headache, but may not reflect the experience of adults with headache who have significant co-morbidities or do not speak English.

All data from included studies was double extracted and agreement checked (KH, TM). However, the COSMIN grading and synthesis score was applied by a single, experienced reviewer (KH). Although applied in several recent reviews [19,84], the grading system itself lacks robust evidence of reliability and validity and should therefore be interpreted with caution.

The lack of reporting guidance and significant heterogeneity in outcome assessment detailed in this review, highlight the importance of establishing guidance on outcome reporting in this population. Future research should seek to establish international, multi-perspective guidance for a core set of outcomes to include in future headache research and across routine practice settings. The first step in this process is to seek consensus on which outcomes should be assessed, as a minimum, in future clinical trials or routine practice settings [85]. Informed by recommendations from this review, the second step is to determine the 'best way' to assess these core outcomes.

Although many PROMs were reviewed following their evaluation in the headache and/or migraine population, study methodological quality was often poor and evidence of essential measurement properties largely unavailable or limited. Such limitations hinder PROM data interpretation from clinical trials, audit or quality assurance initiatives. However, three measures – HIT-6, MSQv2.1 and the PPMQ-R – had acceptable, and often strong, evidence of reliability and validity following completion by patients with headache (HIT-6) or migraine (HIT-6, MSQv2.1, PPMQ-R) and are recommended for consideration in future clinical research and routine practice settings as measures of headache-specific impact, migraine-specific impact, or migraine-treatment response respectively. However, the similarity of item content across all three measures suggests that a further exploration of the attribution, relevance and acceptability of the measures with representative members of the patient population is warranted. Further comparative evidence of widely used generic measures and evidence of measurement responsiveness of all measures is urgently required.

Article Highlights:

- Despite the large number of reviewed PROMs currently used with patients with headache, most have not involved patients in the development process and may lack relevance to the patients experience of headache. Most also lack clarity with regards to measurement aim and have limited evidence of essential measurement properties, limiting confidence in data interpretation. These PROMs should be used and interpreted with caution.
- Strong evidence of reliability and validity was reviewed for three measures – HIT-6, MSQv2.1 and the PPMQ-R – supporting recommendation for consideration in future clinical research or routine practice settings. However, unlike the MSQv2.1 and PPMQ-R, patients were not involved in item generation for the HIT-6.
- The review has highlighted significant heterogeneity in outcome reporting in headache studies, raising concerns over reporting bias and limiting the conduct of systematic reviews and meta-analyses of evidence. International multi-perspective consensus on the most important outcomes – both which outcomes and how to assess – is required, and can be supported by the findings from this review.

Reference list:

1. Lipton RB, Stewart WF, Diamond S, et al. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. *Headache* 2001; 41:646–57.
2. Lipton RB, Bigal ME, Diamond M, et al. Migraine prevalence, disease burden, and the need for preventive therapy. *Neurology* 2007; 68:343–9.
3. Bjorner JB, Kosinski M, Ware JE. Jr. Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HITTM). *Qual Life Res.* 2003;12(8): 913-933.
4. Castillo J, Munoz P, Guitera V, et al. Epidemiology of chronic daily headache in the general population. *Headache* 1999; 39(3): 190–196.
5. Wang S, Fuh J, Lu S, et al. Chronic daily headache in chinese elderly: prevalence, risk factors, and biannual follow-up. *Neurology* 2000; 54: 314–31.
6. Leonardi M, Steiner TJ, Scher AT, et al. The global burden of migraine: Measuring disability in headache disorders with WHO's Classification of Functioning, Disability and Health (ICF). J Headache Pain 2005; 6:429–440.
7. Turner-Bowker DM, Bayliss MS, Ware JE. Jr, Kosinski M. Usefulness of the SF-8TM Health Survey for comparing the impact of migraine and other conditions. *Qual Life Res.* 2003;12(8): 1003-1012.
8. Silberstein S, Tfelt-Hansen P, Dodick DW, et al. Guidelines for controlled trials of prophylactic treatment of chronic migraine in adults. *Cephalalgia.* 2008 May;28(5):484-95.
9. Tfelt-Hansen P, Pascual J, Ramadan N, et al. Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators. *Cephalalgia.* 2012 Jan;32(1):6-38.
10. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A practical guide to their development and use. Oxford University Press. USA. 5th edition. 2014
11. de Vet H, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. A Practical Guide. Cambridge University Press, England. 2011.
12. Rendas-Baum RL, Bloudek M, Maglinte GA, Sepideh F, Varon SF. The psychometric properties of the Migraine-Specific Quality of Life Questionnaire version 2.1 (MSQ) in chronic migraine patients. *Qual Life Res.* 2013; 22(5):1123-1133.
13. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009 Oct;18(8), 1115-1123.
14. Terwee CB, Prinsen CA, Ricci Garotti MG, et al. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res.* 2016 Apr;25(4):767-79.

15. Andrasik F, Lipchik GL, McCrory D, Wittrock DA. Outcome measurement in behavioural headache research: headache parameters and psychosocial outcomes. *Headache.* 2005; 45(5):429-437
16. McCrory DC, Gray RN, Tfelt-Hansen P, Steiner TJ, Taylor FR. Methodological issues in systematic reviews of headache trials: adapting historical diagnostic classifications and outcome measures to present-day standards. *Headache.* 2005; 45(5):459-65.
17. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007 Jan;60(1):34-42.
18. Haywood KL, Staniszewska S, Chapman S. Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *Qual Life Res.* 2012 Feb;21(1):35-52
19. Conijn AP, Jens S, Terwee CB, et al. Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. *Eur J Vasc Endovasc Surg.* 2015 Mar;49(3):316-34
20. Mokkink LB, 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 Jul;63(7):737-45.
21. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012 May;21(4):651-7.
22. Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-reported outcome measures: evolution not revolution. *Patient.* 2012;5(2):79-87
23. Haywood K, Lyddiatt A, Brace-McDonnell SJ, Staniszewska S, Salek S. Establishing the values for patient engagement (PE) in health-related quality of life (HRQoL) research: an international, multiple-stakeholder perspective. *Qual Life Res.* 2016 Dec 8
24. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf.
25. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. *J Nurs Scholarsh.* 2005;37(4):336-42
26. Bakas T, McLennon SM, Carpenter JS, et al. Systematic review of health-related quality of life models. *Health Qual Life Outcomes.* 2012 Nov 16;10:134

27. Elbers RG, Rietberg MB, van Wegen EE, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Qual Life Res.* 2012 Aug;21(6):925-44.
28. Andrée C, Vaillant M, Barre J, et al. Development and validation of the EUROLIGHT questionnaire to evaluate the burden of primary headache disorders in Europe. *Cephalalgia.* 2010 Sep;30(9):1082-100.
29. Vernon H, Lawson G. Development of the headache activities of daily living index: Initial validity study. *J Manipulative Physiol Ther.* 2015; 38(2): 102-111.
30. Niere K, Quin A. Development of a headache-specific disability questionnaire for patients attending physiotherapy. *Man Ther.* 2009 Feb;14(1):45-51
31. Kosinski M, Bayliss MS, Bjorner JB, et al. A six-item short-form survey for measuring headache impact: The HIT-6. *Qual Life Res.* 2003; 12(8): 963-974.
32. Magnusson JE, Riess CM, Becker WJ. Modification of the SF-36 for a headache population changes patient-reported health status. *Headache.* 2012; 52(6): 993-1004.
33. Pathak DS, Chisolm DJ, Weis KA. Functional Assessment in Migraine (FAIM) questionnaire: development of an instrument based upon the WHO's International Classification of Functioning, Disability, and Health. *Value Health*. 2005; 8(5):591-600.
34. Cramer JA, Silberstein SD, Winner P. Development and validation of the headache needs assessment (HANA) survey. *Headache.* 2001; 41(4): 402-409.
35. Stewart WF, Lipton RB, Whyte J, et al. An international study to assess reliability of the Migraine Disability Assessment (MIDAS) score. Neurology. 1999 22;53(5):988-94.
36. Martin BC, Pathak DS, Sharfman MI, et al. Validity and reliability of the migraine-specific quality of life questionnaire (MSQ Version 2.1). *Headache.* 2000; 40(3): 204-215.
37. McKenna SP, Doward LC, Davey KM. The Development and Psychometric Properties of the MSQOL: A Migraine-Specific Quality-of-Life Instrument. *Clin Drug Investig.* 1998;15(5):413-23.
38. Coon CD, Fehnel SE, Davis KH, et al. The development of a survey to measure completeness of response to migraine therapy. *Headache.* 2012; 52(4): 550-572.
39. Dowson AJ, Tepper SJ, Baos V, et al. Identifying patients who require a change in their current acute migraine treatment: The Migraine Assessment of Current Therapy (Migraine-ACT) questionnaire. *Curr Med Res Opin.* 2004; 20(7): 1125-1135.
40. Chatterton ML, Shechter A, Curtice WS, et al. Reliability and validity of the migraine therapy assessment questionnaire. *Headache.* 2002; 42(10): 1006-1015.
41. Lipton RB, Kolodner K, Bigal ME, et al. Validity and reliability of the migraine-treatment

optimization questionnaire. *Cephalalgia.* 2009; 29(7): 751-759.

42. Patrick DL, Martin ML, Bushnell DM, Pesa J. Measuring satisfaction with migraine treatment: expectations, importance, outcomes, and global ratings. *Clin Ther.* 2003; 25(11): 2920-2935.
43. Revicki DA, Kimel M, Beusterien K, et al. Validation of the revised patient perception of migraine questionnaire: Measuring satisfaction with acute migraine treatment. *Headache.* 2006; 46(2): 240-252.
44. Ware JE, Sherbourne CD. The MOS 36-item Short Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30:473-83.
45. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34(3): 220-33.
46. Ware J, Kosinski M, Dewey J, Gandek B: How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey. Boston: Quality Metric; 2001.
47. The EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. Health Policy. 1990;16: 199-208.
48. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care.* 2002 Feb;40(2):113-28.
49. Kaplan RM. Quality of life assessment for cost/utility studies in cancer. *Cancer Treat Rev.* 1993;19 Suppl A:85-96.
50. Andresen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Med Care.* 1998 Sep;36(9):1349-60.
51. Cole JC, Lin P, Rupnow MFT. Validation of the Migraine-Specific Quality of Life Questionnaire version 2.1 (MSQ v. 2.1) for patients undergoing prophylactic migraine treatment. *Qual Life Res.* 2007; 16(7): 1231-1237.
52. Patrick DL, Hurst BC, Hughes J. Further development and testing of the migraine-specific quality of life (MSQOL) measure. *Headache.* 2000; 40(7): 550-560.
53. Rendas-Baum R, Yang M, Varon SF, et al. Validation of the Headache Impact Test (HIT-6) in patients with chronic migraine. *Health Qual Life Outcomes.* 2014; 12: 117.
54. Bagley CL, Rendas-Baum R, Maglinte GA, et al. Validating migraine-specific quality of life questionnaire v2.1 in episodic and chronic migraine. *Headache.* 2012; 52(3): 409-421.
55. Kawata AK, Coeytaux RR, DeVellis RF, et al. Psychometric properties of the HIT-6 among patients in a headache-specialty practice. *Headache.* 2005; 45(6): 638-643.
56. Yang M, Rendas-Baum R, Varon SF, Kosinski M. Validation of the Headache Impact Test (HIT-

6TM) across episodic and chronic migraine. *Cephalalgia.* 2011; 31(3): 357-367.

57. Ware JE Jr, Kosinski M, Bjorner JB, et al. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res.* 2003 Dec;12(8):935-52.
58. Coeytaux RR, Kaufman JS, Chao R, et al. Four methods of estimating the minimal important difference score were compared to establish a clinically significant change in Headache Impact Test. *J Clin Epidemiol.* 2006; 59(4): 374-380.
59. Lipton RB, Rosen NL, Ailani J et al. Pooled results from the PREEMPT randomized clinical trial program. *Cephalalgia.* 2016; 36(9): 899-908.
60. Cole J, Lin P, Rupnow M. Minimal important differences in the Migraine-Specific Quality of Life Questionnaire (MSQ) version 2.1. *Cephalalgia.* 2009; 29(11): 1180-1187.
61. Dodick DW, Silberstein S, Saper J et al. The impact of topiramate on health-related quality of life indicators in chronic migraine. *Headache.* 2007; 47: 1398-1408.
62. Castien RF, Blankenstein AH, Windt DA, Dekker J. Minimal clinically important change on the Headache Impact Test-6 questionnaire in patients with chronic tension-type headache. *Cephalalgia.* 2012 Jul;32(9):710-4.
63. Smelt AF, Assendelft WJ, Terwee CB, Ferrari MD, Blom JW. What is a clinically relevant change on the HIT-6 questionnaire? An estimation in a primary-care population of migraine patients. *Cephalalgia.* 2014 Jan;34(1):29-36.
64. Sieber WJ, David KM, Adams JE, et al. Assessing the impact of migraine on health-related quality of life: An additional use of the quality of well-being scale-self-administered. *Headache.* 2000; 40(8): 662-671.
65. Martin ML, Patrick DL, Bushnell DM, et al. Further validation of an individualized migraine treatment satisfaction measure. *Value Health*. 2008; 11(5): 904-912.
66. Stewart WF, Lipton R, Kolodner K, et al. Reliability of the migraine disability assessment score in a population- based sample of headache sufferers. *Cephalalgia.* 1999; 19(2): 107-114.
67. Stewart WF, Lipton RB, Kolodner KB, et al. Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain.* 2000; 88(1): 41-52.
68. Stewart WF, Lipton RB, Kolodner K. Migraine disability assessment (MIDAS) score: relation to headache frequency, pain intensity, and headache symptoms. *Headache.* 2003 Mar;43(3):258-65.
69. Bigal ME, Rapoport AM, Lipton RB, et al. Assessment of migraine disability using the migraine disability assessment (MIDAS) questionnaire: a comparison of chronic migraine with episodic migraine. *Headache.* 2003; 43(4): 336-342.

70. Sauro KM, Rose MS, Becker WJ, et al. HIT-6 and MIDAS as measures of headache disability in a headache referral population. *Headache.* 2010 Mar;50(3):383-95
71. Blumenfeld AM, Varon SF, Wilcox TK, et al. Disability, HRQoL and resource use among chronic and episodic migraineurs: results from the International Burden of Migraine Study (IBMS). *Cephalalgia*. 2011; 31(3): 301-315.
72. Stafford MR, Hareendran A, Ng-Mak DS, et al. EQ-5D-derived utility values for different levels of migraine severity from a UK sample of migraineurs. Health and Quality of Life Outcomes 2012 Jun; 10(65).
73. Gillard PJ, Devine B, Varon SF, et al. Mapping from disease-specific measures to health-state utility values in individuals with migraine. *Value Health.* 2012; 15(3): 485-494.
74. Kilminster SG, Dowson AJ, Tepper SJ, et al. Reliability, validity, and clinical utility of the migraine-ACT questionnaire. *Headache.* 2006; 46(4): 553-562.
75. Kimel M, Hsieh R, McCormack SP, et al. Validation of the revised Patient Perception of Migraine Questionnaire (PPMQ-R): measuring satisfaction with acute migraine treatment in clinical trials. *Cephalalgia.* 2008; 28(5): 510-523.
76. Davis KH, Black L, Sleath B. Validation of the Patient Perception of Migraine Questionnaire. *Value Health.* 2002; 5(5): 422-430.
77. Lipton RB, Bigal ME, Stewart WF. Assessing disability using the migraine disability assessment questionnaire. *Expert Rev Neurother*. 2003; 3(3): 317-325.
78. Xu R, Insinga RP, Golden W, Hu XH. EuroQol (EQ-5D) health utility scores for patients with migraine. *Qual Life Res.* 2011; 20(4): 601-608.
79. Brown JS, Neumann PJ, Papadopoulos G, et al. Migraine frequency and health utilities: findings from a multisite survey. *Value Health.* 2008; 11(2):315-21
80. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity-establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1--eliciting concepts for a new PRO instrument. *Value Health*. 2011 Dec;14(8):967-77
81. Blau JN. Fears aroused in patients by migraine. *Br Med J (Clin Res).* 1984; 288:1126.
82. Wagner TH, Patrick DL, Galer BS, Berzon RA. A new instrument to assess the long-term quality of life effects from migraine: development and psychometric testing of the MSQOL. *Headache.* 1996 Sep;36(8):484-92.
83. Yost KJ, Cella D, Chawla A, et al. Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a

combination of distribution- and anchor-based approaches. *J Clin Epidemiol.* 2005 Dec;58(12):1241-51

84. Haywood KL, Brett J, Tutton E, Staniszewska S. Patient-reported outcome measures in older people with hip fracture: a systematic review of quality and acceptability. *Qual Life Res.* 2016 Oct 20. [Epub ahead of print]
85. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012 Aug 6;13:132

Tables and Figures:

Figure 1: Review of measures used with people with headache – PRISMA flow diagram for article selection (search conducted 1980 to December 2016).

Table 1: Data synthesis, levels of evidence and overall quality of reviewed PROMs in the headache population (n=23)

Appendices:

Appendix 1: Search strategies – including named measure searches.

Appendix 2: Table 1: Multi-item PROMs identified (n=39) from full-text articles assessed for eligibility (searched 1980-2016); n=23 PROMs included in final review ('in-use' 2000-2016).

Appendix 3: Table 2: Assessment criteria for the quality of reported measurement properties [17,19,20].

Appendix 4: Table 3: Characteristics of reviewed PROMs evaluated in the headache population (total = 23)

Appendix 5: Table 4: Characteristics of included studies (n= 46)

Appendix 6: Table 5: Content comparison at item level (number of items) of condition-specific measures (n= 17)

Appendix 7: Table 6: Methodological quality (COSMIN[a]) of each study (n=46) per PROM (n=23) and investigated measurement property.

Funding source:

This research was funded by the NIHR Programme Grants for Applied Research programme (RP-PG-1212-20018). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Authors contribution:

KH, MU and MM conceived and designed the study; KH organized the conduct of the study; TM and KH carried out the study (including acquisition of study data); TM, KH, RP and SP contributed to search strategy design and data extraction; KH and TM double extracted all data; KH completed the analysis and interpretation of study data. KH and TM drafted the output; all authors provided a critique of the output for important intellectual content.

Acknowledgements: The CHESS team: Professor Martin Underwood (Chief investigator), Felix Achana, David Boss, Ms Mary Bright, Fiona Caldwell, Dr Dawn Carnes, Dr Brendan Davies, Professor Sandra Eldridge, Dr David Ellard, Simon Evans, Professor Frances Griffiths, Dr Kirstie Haywood, Dr Siew Wan Hee, Dr Manjit Matharu, Hema Mistry, Professor Stavros Petrou, Professor Tamar Pincus, Dr Katrin Probyn, Dr Harbinder Sandhu, Professor Stephanie Taylor, Arlene Wilkie, Helen Higgins, Dr Vivien Nichols, Dr Shilpa Patel, Dr Rachel Potter, Kimberley White.

Headache Review – Cephalalgia

Appendix 1.1 : Search strategies: Construct searches

Database: Embase Classic+Embase <1947 to 2017 Week 16>

Search Strategy:

---

1 exp chronic daily headache/ or exp episodic tension headache/ or exp headache/ or exp primary headache/ or exp chronic tension headache/ or exp new daily persistent headache/ or exp secondary headache/ or exp tension headache/ (201706)

2 (headache* or migraine*).ti,ab. (136921)

3 (headache* adj3 (mixed or combination or tension or tension type or muscle contraction or psychomyogenic or stress or ordinary or essential or psychogenic)).tw. (6527)

4 ((chronic adj2 daily adj2 headache*) or (daily adj2 persistent adj2 headache*)).ti,ab. (1530)

5 long term headache*.tw. (43)

6 1 or 2 or 3 or 4 or 5 (244190)

7 (daily or persistent or chronic).mp. (2456082)

8 6 and 7 (56510)

9 exp migraine aura/ or exp ophthalmoplegic migraine/ or exp migraine/ or exp migraine with aura/ or exp migraine without aura/ (57058)

10 ((withdrawal or overuse or "over use" or "over-use" or misuse or "mis-use" or abuse or induced) adj5 (medication* or medicine* or analges* or drug* or opiate* or opioid* or nsaids or non-opiate* or non opiate or ergot* or painkiller* or pain killer* or pain-killer*) adj5 (headache* or migraine*)).mp. (20637)

11 ((rebound or transformed) adj5 (headache* or migraine*)).ti,ab. (450)

12 8 or 9 or 10 or 11 (119498)

13 (HR-PRO or HRPRO or HRPRO or HRQL or HRQoL or QL or QoL).mp. (73860)

14 (quality of life or life quality).tw. (310350)

15 (health index or health indices or health profile).mp. (5057)

16 (patient or self or child or parent or carer or proxy).mp. (8459912)

17 (report or reported or reporting or rated or rating or ratings or based or assessed or assessment assessments or disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being).mp. (12222752)

18 (index or indices or instrument or instruments or measure or measures or questionnaire or questionnaires or profile or profiles or scale or scales or score or scores or status or survey or surveys).ti,ab. (5157109)

19 health related quality of life.ti,ab. (44869)

20 quality adjusted life year.ti,ab. (5178)

21 QALY.tw. (11466)

22 value of life.tw. (309)

23 ((health adj2 utility*) or disutili*).mp. (2799)

24 willingness to pay.tw. (5728)

25 contingent valuation.tw. (697)

26 standard gamble.tw. (945)

27 SG.tw. (11070)

28 time tradeoff.tw. (250)

29 time trade off.tw. (1369)

30 TTO.tw. (1299)

31 mapping.tw. (161946)

32 cross walking.tw. (17)

33 transfer to utility.tw. (12)

34 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 (17788699)

35 exp 'intermethod comparison'/ or exp 'data collection method'/ or exp 'validation study'/ or exp 'feasibility study'/ or exp 'pilot study'/ or exp 'psychometry'/ or exp 'reproducibility'/ or reproducib*.ab,ti. or 'audit'.ab,ti. or psychometr*.ab,ti. or clinimetr*.ab,ti. or clinometr*.ab,ti. or exp 'observer variation'/ or 'observer variation'.ab,ti. or exp 'discriminant analysis'/ or exp 'validity'/ or reliab*.ab,ti. or valid*.ab,ti. or 'coefficient'.ab,ti. or 'internal consistency'.ab,ti. or (cronbach* and ('alpha' or 'alphas')).ab,ti. or 'item correlation'.ab,ti. or 'item correlations'.ab,ti. or 'item selection'.ab,ti. or 'item selections'.ab,ti. or 'item reduction'.ab,ti. or 'item reductions'.ab,ti. or 'agreement'.ab,ti. or 'precision'.ab,ti. or 'imprecision'.ab,ti. or 'precise values'.ab,ti. or 'test-retest'.ab,ti. or ('test' and 'retest').ab,ti. or (reliab* and ('test' or 'retest')).ab,ti. or 'stability'.ab,ti. or 'interrater'.ab,ti. or 'inter-rater'.ab,ti. or 'intrarater'.ab,ti. or 'intra-rater'.ab,ti. or 'intertester'.ab,ti. or 'inter-tester'.ab,ti. or 'intratester'.ab,ti. or 'intra- tester'.ab,ti. or 'interobeserver'.ab,ti. or 'inter-observer'.ab,ti. or 'intraobserver'.ab,ti. or 'intra- observer'.ab,ti. or 'intertechnician'.ab,ti. or 'inter-technician'.ab,ti. or 'intratechnician'.ab,ti. or 'intra- technician'.ab,ti. or 'interexaminer'.ab,ti. or 'inter-examiner'.ab,ti. or 'intraexaminer'.ab,ti. or 'intra-examiner'.ab,ti. or 'interassay'.ab,ti. or 'inter-assay'.ab,ti. or 'intraassay'.ab,ti. or 'intra-assay'.ab,ti. or 'interindividual'.ab,ti. or 'inter-individual'.ab,ti. or 'intraindividual'.ab,ti. or 'intra-individual'.ab,ti. or 'interparticipant'.ab,ti. or 'inter-participant'.ab,ti. or 'intraparticipant'.ab,ti. or 'intra- participant'.ab,ti. or 'kappa'.ab,ti. or 'kappas'.ab,ti. or 'coefficient of variation'.ab,ti. or repeatab*.ab,ti. or ((replicab* or 'repeated') and ('measure' or 'measures' or 'findings' or 'result' or 'results' or 'test' or 'tests')).ab,ti. or generaliza*.ab,ti. or generalisa*.ab,ti. or 'concordance'.ab,ti. or ('intraclass' and correlation*).ab,ti. or 'discriminative'.ab,ti. or 'known group'.ab,ti. or 'factor analysis'.ab,ti. or 'factor analyses'.ab,ti. or 'factor structure'.ab,ti. or 'factor structures'.ab,ti. or

'dimensionality'.ab,ti. or subscale*.ab,ti. or 'multitrait scaling analysis'.ab,ti. or 'multitrait scaling analyses'.ab,ti. or 'item discriminant'.ab,ti. or 'interscale correlation'.ab,ti. or 'interscale correlations'.ab,ti. or (('error' or 'errors') and (measure* or correlat* or evaluat* or 'accuracy' or 'accurate' or 'precision' or 'mean')).ab,ti. or 'individual variability'.ab,ti. or 'interval variability'.ab,ti. or 'rate variability'.ab,ti. or 'variability analysis'.ab,ti. or ('uncertainty' and ('measurement' or 'measuring')).ab,ti. or 'standard error of measurement'.ab,ti. or sensitiv*.ab,ti. or responsive*.ab,ti. or ('limit' and 'detection').ab,ti. or 'minimal detectable concentration'.ab,ti. or interpretab*.ab,ti. or (small* and ('real' or 'detectable') and ('change' or 'difference')).ab,ti. or 'meaningful change'.ab,ti. or 'minimal important change'.ab,ti. or 'minimal important difference'.ab,ti. or 'minimally important change'.ab,ti. or 'minimally important difference'.ab,ti. or 'minimal detectable change'.ab,ti. or 'minimal detectable difference'.ab,ti. or 'minimally detectable change'.ab,ti. or 'minimally detectable difference'.ab,ti. or 'minimal real change'.ab,ti. or 'minimal real difference'.ab,ti. or 'minimally real change'.ab,ti. or 'minimally real difference'.ab,ti. or 'ceiling effect'.ab,ti. or 'floor effect'.ab,ti. or 'item response model'.ab,ti. or 'irt'.ab,ti. or 'rasch'.ab,ti. or 'differential item functioning'.ab,ti. or 'dif'.ab,ti. or 'computer adaptive testing'.ab,ti. or 'item bank'.ab,ti. or 'cross-cultural equivalence'.ab,ti. (5156030)

36 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (1526497)

37 (12 and 34 and 35) not 36 (20124)

38 limit 37 to (human and english language and (adult <18 to 64 years> or aged <65+ years>)) (10502)

39 limit 38 to yr="1980 - 2016" (10296)

Database: Ovid MEDLINE(R) <1946 to April Week 2 2017>

Search Strategy:

---

1 Headache/ (26018)

2 exp headache disorders/ or exp headache disorders, primary/ (32133)

3 exp Tension-Type Headache/ (1860)

4 (headache* adj3 (mixed or combination or tension or tension type or muscle contraction* or psychomyogenic or stress or ordinary or essential or psychogenic)).tw. (4133)

5 ((chronic adj2 daily adj2 headache*) or (daily adj2 persistent adj2 headache*)).ti,ab. (950)

6 (headache* or hemicrania simplex).mp. (74687)

7 long term headache*.mp. (19)

8 chronic headache*.mp. (1518)

9 exp Headache Disorders/ (32133)

10 tension headache*.mp. (1004)

11 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (88376)

12 (daily or persistent or chronic).mp. (1583660)

13 11 and 12 (17326)

14 migraine*.mp. or exp Migraine with Aura/ or exp Migraine Disorders/ or exp Ophthalmoplegic Migraine/ or exp Migraine without Aura/ (32683)

15 (withdrawal or overuse or "over use" or "over-use" or misuse or "mis-use" or abuse or induced).mp. adj5 (medication* or medicine* or analges* or drug* or opiate* or opioid* or NSAIDS or non-opiate* or non opiate* or ergot* or painkiller* or pain killer* or pain-killer*).ti,ab. adj5 (headache* or migraine*).ti,ab. (3474)

16 ((rebound or transformed) adj5 (headache* or migraine*)).ti,ab. (327)

17 13 or 14 or 15 or 16 (45932)

18 (HR-PRO or HRPRO or HRPRO or HRQL or HRQoL or QL or QoL or PRO or PROs or PROM or PROMs).mp. (189311)

19 (quality of life or life quality).mp. (228883)

20 (health index* or health indices or health profile* or health status).mp. (128028)

21 ((patient or self or proxy) adj (appraisal* or appraised or report or reported or reporting or rated or rating or ratings or based or assessed or assessment*)).mp. (155466)

22 ((disability or function or functional or functions or subjective or utility or utilities or wellbeing or well being or health) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile or profiles or scale or scales or score or scores or status or survey or surveys)).ti,ab. (182045)

23 health related quality of life.ti,ab. (27920)

24 quality adjusted life year.ti,ab. (3358)

25 QALY.tw. (5322)

26 value of life.tw. (249)

27 ((health adj2 utility*) or disutili*).mp. (1384)

28 willingness to pay.tw. (3086)

29 contingent valuation.tw. (498)

30 standard gamble.tw. (749)

31 SG.tw. (6571)

32 time tradeoff.tw. (252)

33 time trade off.tw. (884)

34 TTO.tw. (746)

35 mapping.tw. (122495)

36 cross walking.tw. (6)

37 transfer to utility.tw. (8)

38 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 (834011)

39 (instrumentation or methods).sh. or Validation Stud- ies.pt. or Comparative Study.pt. or exp Psychometrics/ or psychometr*.ti,ab. or clinimetr*.tw. or clino- metr*.tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti,ab. or outcome measure*.tw. or exp observer variation/ or observer variation.ti,ab. or exp Health Status Indicators/ or exp reproducibility of results/ or reproducib*.ti,ab. or exp discriminant analysis/ or reliab*.ti,ab. or unreliab*.ti,ab. or valid*.ti,ab. or coefficient.ti,ab. or homogeneity.ti,ab. or homogeneous.ti,ab. or internal consistency.ti,ab. or (cronbach* and (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or agreement.ti,ab. or precision.ti,ab. or imprecision.ti,ab. or "precise values".ti,ab. or test- retest.ti,ab. or (test and retest).ti,ab. or (reliab* and (test or retest)).ti,ab. or stability.ti,ab. or interrater.ti,ab. or inter-rater.ti,ab. or intrarater.ti,ab. or intra-rater.ti,ab. or intertester.ti,ab. or inter-tester.ti,ab. or intratester.ti,ab. or intra-tester.ti,ab. or interobserver.ti,ab. or inter-observer.ti,ab. or intraobserver.ti,ab. or intra- observer.ti,ab. or intertechnician.ti,ab. or inter-techni- cian.ti,ab. or intratechnician.ti,ab. or intra-technician.ti,ab. or interexaminer.ti,ab. or inter-examiner.ti,ab. or intraex- aminer.ti,ab. or intra-examiner.ti,ab. or interassay.ti,ab. or inter-assay.ti,ab. or intraassay.ti,ab. or intra-assay.ti,ab. or interindividual.ti,ab. or inter-individual.ti,ab. or intraindi- vidual.ti,ab. or intra-individual.ti,ab. or interparticipant.ti,ab. or inter-participant.ti,ab. or intraparticipant.ti,ab. or intra-participant.ti,ab. or kappa.ti,ab. or kappa-s.ti,ab. or kappas.ti,ab. or repeatab*.ti,ab. or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. or generaliza*.ti,ab. or general- isa*.ti,ab. or concordance.ti,ab. or (intraclass and correlation*).ti,ab. or discriminative.ti,ab. or "known group".ti,ab. or factor analysis.ti,ab. or factor analyses.ti,ab. or dimension*.ti,ab. or subscale*.ti,ab. or (multitrait and scaling and (analysis or analyses)).ti,ab. or item discriminant.ti,ab. or interscale correlation*.ti,ab. or

error.ti,ab. or errors.ti,ab. or "individual variabil- ity".ti,ab. or (variability and (analysis or val- ues)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or "standard error of measure-ment".ti,ab. or sensitiv*.ti,ab. or responsive*.ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. or (small* and (real or detectable) and (change or difference)).ti,ab. or meaningful change.ti,ab. or "ceiling effect".ti,ab. or "floor effect".ti,ab. or "Item response model".ti,ab. or IRT.ti,ab. or Rasch.ti,ab. or "Differential item functioning".ti,ab. or DIF.ti,ab. or "computer adaptive testing".ti,ab. or "item bank".ti,ab. or "cross-cultural equivalence".ti,ab. (5772794)

40 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (3636046)

41 (17 and 38 and 39) not 40 (2028)

42 limit 41 to (english language and humans and "all adult (19 plus years)") (1412)

43 limit 42 to yr="1980 - 2016" (1405)

Appendix 1.2 Named measure searches

Database: Embase Classic+Embase <1947 to 2017 Week 16>

Search Strategy:

---

1 exp chronic daily headache/ or exp episodic tension headache/ or exp headache/ or exp primary headache/ or exp chronic tension headache/ or exp new daily persistent headache/ or exp secondary headache/ or exp tension headache/ (201706)

2 (headache* or migraine*).ti,ab. (136921)

3 (headache* adj3 (mixed or combination or tension or tension type or muscle contraction or psychomyogenic or stress or ordinary or essential or psychogenic)).tw. (6527)

4 ((chronic adj2 daily adj2 headache*) or (daily adj2 persistent adj2 headache*)).ti,ab. (1530)

5 long term headache*.tw. (43)

6 1 or 2 or 3 or 4 or 5 (244190)

7 (daily or persistent or chronic).mp. (2456082)

8 6 and 7 (56510)

9 exp migraine aura/ or exp ophthalmoplegic migraine/ or exp migraine/ or exp migraine with aura/ or exp migraine without aura/ (57058)

10 ((withdrawal or overuse or "over use" or "over-use" or misuse or "mis-use" or abuse or induced) adj5 (medication* or medicine* or analges* or drug* or opiate* or opioid* or nsaids or non-opiate* or non opiate or ergot* or painkiller* or pain killer* or pain-killer*) adj5 (headache* or migraine*)).mp. (20637)

11 ((rebound or transformed) adj5 (headache* or migraine*)).ti,ab. (450)

12 8 or 9 or 10 or 11 (119498)

13 exp 'intermethod comparison'/ or exp 'data collection method'/ or exp 'validation study'/ or exp 'feasibility study'/ or exp 'pilot study'/ or exp 'psychometry'/ or exp 'reproducibility'/ or reproducib*.ab,ti. or 'audit'.ab,ti. or psychometr*.ab,ti. or clinimetr*.ab,ti. or clinometr*.ab,ti. or exp 'observer variation'/ or 'observer variation'.ab,ti. or exp 'discriminant analysis'/ or exp 'validity'/ or reliab*.ab,ti. or valid*.ab,ti. or 'coefficient'.ab,ti. or 'internal consistency'.ab,ti. or (cronbach* and ('alpha' or 'alphas')).ab,ti. or 'item correlation'.ab,ti. or 'item correlations'.ab,ti. or 'item selection'.ab,ti. or 'item selections'.ab,ti. or 'item reduction'.ab,ti. or 'item reductions'.ab,ti. or 'agreement'.ab,ti. or 'precision'.ab,ti. or 'imprecision'.ab,ti. or 'precise values'.ab,ti. or 'test-retest'.ab,ti. or ('test' and 'retest').ab,ti. or (reliab* and ('test' or 'retest')).ab,ti. or 'stability'.ab,ti. or 'interrater'.ab,ti. or 'inter-rater'.ab,ti. or 'intrarater'.ab,ti. or 'intra-rater'.ab,ti. or 'intertester'.ab,ti. or 'inter-tester'.ab,ti. or 'intratester'.ab,ti. or 'intra- tester'.ab,ti. or 'interobeserver'.ab,ti. or 'inter-observer'.ab,ti. or 'intraobserver'.ab,ti. or 'intra- observer'.ab,ti. or 'intertechnician'.ab,ti. or 'inter-technician'.ab,ti. or 'intratechnician'.ab,ti. or 'intra- technician'.ab,ti. or 'interexaminer'.ab,ti. or 'inter-examiner'.ab,ti. or 'intraexaminer'.ab,ti. or 'intra-examiner'.ab,ti. or 'interassay'.ab,ti. or 'inter-assay'.ab,ti. or 'intraassay'.ab,ti. or 'intra-assay'.ab,ti. or 'interindividual'.ab,ti. or 'inter-individual'.ab,ti. or 'intraindividual'.ab,ti. or 'intra-individual'.ab,ti. or 'interparticipant'.ab,ti. or 'inter-participant'.ab,ti. or 'intraparticipant'.ab,ti. or 'intra- participant'.ab,ti. or 'kappa'.ab,ti. or 'kappas'.ab,ti. or 'coefficient of variation'.ab,ti. or repeatab*.ab,ti. or ((replicab* or 'repeated') and ('measure' or 'measures' or 'findings' or 'result' or 'results' or 'test' or 'tests')).ab,ti. or generaliza*.ab,ti. or generalisa*.ab,ti. or 'concordance'.ab,ti. or ('intraclass' and correlation*).ab,ti. or 'discriminative'.ab,ti. or 'known group'.ab,ti. or 'factor analysis'.ab,ti. or 'factor analyses'.ab,ti. or 'factor structure'.ab,ti. or 'factor structures'.ab,ti. or 'dimensionality'.ab,ti. or subscale*.ab,ti. or 'multitrait scaling analysis'.ab,ti. or 'multitrait scaling analyses'.ab,ti. or 'item discriminant'.ab,ti. or 'interscale correlation'.ab,ti. or 'interscale

correlations'.ab,ti. or (('error' or 'errors') and (measure* or correlat* or evaluat* or 'accuracy' or 'accurate' or 'precision' or 'mean')).ab,ti. or 'individual variability'.ab,ti. or 'interval variability'.ab,ti. or 'rate variability'.ab,ti. or 'variability analysis'.ab,ti. or ('uncertainty' and ('measurement' or 'measuring')).ab,ti. or 'standard error of measurement'.ab,ti. or sensitiv*.ab,ti. or responsive*.ab,ti. or ('limit' and 'detection').ab,ti. or 'minimal detectable concentration'.ab,ti. or interpretab*.ab,ti. or (small* and ('real' or 'detectable') and ('change' or 'difference')).ab,ti. or 'meaningful change'.ab,ti. or 'minimal important change'.ab,ti. or 'minimal important difference'.ab,ti. or 'minimally important change'.ab,ti. or 'minimally important difference'.ab,ti. or 'minimal detectable change'.ab,ti. or 'minimal detectable difference'.ab,ti. or 'minimally detectable change'.ab,ti. or 'minimally detectable difference'.ab,ti. or 'minimal real change'.ab,ti. or 'minimal real difference'.ab,ti. or 'minimally real change'.ab,ti. or 'minimally real difference'.ab,ti. or 'ceiling effect'.ab,ti. or 'floor effect'.ab,ti. or 'item response model'.ab,ti. or 'irt'.ab,ti. or 'rasch'.ab,ti. or 'differential item functioning'.ab,ti. or 'dif'.ab,ti. or 'computer adaptive testing'.ab,ti. or 'item bank'.ab,ti. or 'cross-cultural equivalence'.ab,ti. (5156030)

14 (burden of migraine questionnaire or BURMIG or (beck depression inventory or BDI) or (comprehensive headache related quality of life questionnaire or CHQQ) or (completeness of response survey or CORS) or (cognitive impairment scale for migraine attacks or MIG-SCOG) or (chronic pain coping inventory or CPCI) or (depression anxiety stress scale or DASS) or (functional assessment in migraine or FAIM or FAIMQ) or (female sexual function index or FSFI) or (health utilities index or HUI) or (headache impact test or HIT 6 or HIT6 or HIT-6) or (headache impact score or HIS) or (headache management self-efficacy scale or HSES) or (headache-specific locus of control or headache specific locus of control) or (headache disability inventory or HDI) or (headache disability scale or HDS) or ((hospital anxiety and depression scale) or HADS) or ((headache-attributed restriction, disability, social handicap and impaired participation questionnaire) or HARDSHIP or HARDSHIPQ) or (headache activities of daily living index or HADLI) or (headache under response to treatment questionnaire or HURT or HURTQ) or (headache impact questionnaire or HIQ or HImQ) or (headache needs assessment survey or HANA) or (headache intensity or headache duration or headache severity) or (henry ford hospital headache disability inventory or HDI) or (impact of migraine-tension type headache-neck pain or Impact M-TTH-NP) or (italian perceived disability scale or IPDS) or (migraine treatment optimisation questionnaire or M-TOQ-15) or (migraine treatment satisfaction measure or MTSM) or (migraine-specific quality of life scale or MSQoL) or ((migraine work and productivity loss questionnaire) or MWPLQ) or (migraine disability assessment score or MIDAS) or (migraine disability assessment questionnaire or MDAS) or (migraine screen questionnaire or MS-Q) or (migraine impact questionnaire or MIQ) or (migraine specific quality of life questionnaire or MSQL or MSQ or MSQV 2 1) or (24-h MSQoLQ or 24-hour migraine specific quality of life questionnaire) or (patient perception of migraine questionnaire or PPMQ) or (patient health questionnaire or PHQ-9) or (pain disability index or PDI) or (pittsburg sleep quality index or PSQI) or (subjects global impression of change or SGIC) or (pain catastrophizing scale or PCS) or (visual aura rating scale or VARS) or (waters headache questionnaire or WHQ) or (numerical rating scale or NRS or numerical pain intensity scale or numerical pain rating scale or numeric rating scale for pain or NRS pain or NRS-pain) or (visual analogue scale or visual analogue scale for pain or VAS pain or VAS-pain or VAS) or (rating scale or analogue scale) or (SF36 or SF 36 or SF-36 or short form 36 or shortform 36 or short-form 36) or (SF12 or SF 12 or SF-12 or short form 12 or shortform 12 or short-form 12) or (SF6D or SF 6D or SF-6D or short form 6D or shortform 6D or short-form 6D) or (euroqol or euro qol or euro-qol or EQ5D or EQ 5D or EQ-5D)).tw. (552600)

15 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (1526497)

16 (12 and 13 and 14) not 15 (3250)

17 limit 16 to (human and english language and (adult <18 to 64 years> or aged <65+ years>)) (1794)

18 limit 17 to yr="1980 - 2016" (1760)

Database: Ovid MEDLINE(R) <1946 to April Week 2 2017>

Search Strategy:

---

1 Headache/ (26018)

2 exp headache disorders/ or exp headache disorders, primary/ (32133)

3 exp Tension-Type Headache/ (1860)

4 (headache* adj3 (mixed or combination or tension or tension type or muscle contraction* or psychomyogenic or stress or ordinary or essential or psychogenic)).tw. (4133)

5 ((chronic adj2 daily adj2 headache*) or (daily adj2 persistent adj2 headache*)).ti,ab. (950)

6 (headache* or hemicrania simplex).mp. (74687)

7 long term headache*.mp. (19)

8 chronic headache*.mp. (1518)

9 exp Headache Disorders/ (32133)

10 tension headache*.mp. (1004)

11 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (88376)

12 (daily or persistent or chronic).mp. (1583660)

13 11 and 12 (17326)

14 migraine*.mp. or exp Migraine with Aura/ or exp Migraine Disorders/ or exp Ophthalmoplegic Migraine/ or exp Migraine without Aura/ (32683)

15 (withdrawal or overuse or "over use" or "over-use" or misuse or "mis-use" or abuse or induced).mp. adj5 (medication* or medicine* or analges* or drug* or opiate* or opioid* or NSAIDS or non-opiate* or non opiate* or ergot* or painkiller* or pain killer* or pain-killer*).ti,ab. adj5 (headache* or migraine*).ti,ab. (3474)

16 ((rebound or transformed) adj5 (headache* or migraine*)).ti,ab. (327)

17 13 or 14 or 15 or 16 (45932)

18 (instrumentation or methods).sh. or Validation Stud- ies.pt. or Comparative Study.pt. or exp Psychometrics/ or psychometr*.ti,ab. or clinimetr*.tw. or clino- metr*.tw. or exp "Outcome Assessment (Health Care)"/ or outcome assessment.ti,ab. or outcome measure*.tw. or exp observer variation/ or observer variation.ti,ab. or exp Health Status Indicators/ or exp reproducibility of results/ or reproducib*.ti,ab. or exp discriminant analysis/ or reliab*.ti,ab. or unreliab*.ti,ab. or valid*.ti,ab. or coefficient.ti,ab. or homogeneity.ti,ab. or homogeneous.ti,ab. or internal consistency.ti,ab. or (cronbach* and (alpha or alphas)).ti,ab. or (item and (correlation* or selection* or reduction*)).ti,ab. or agreement.ti,ab. or precision.ti,ab. or imprecision.ti,ab. or "precise values".ti,ab. or test- retest.ti,ab. or (test and retest).ti,ab. or (reliab* and (test or retest)).ti,ab. or stability.ti,ab. or interrater.ti,ab. or inter-rater.ti,ab. or intrarater.ti,ab. or intra-rater.ti,ab. or intertester.ti,ab. or inter-tester.ti,ab. or intratester.ti,ab. or intra-tester.ti,ab. or interobserver.ti,ab. or inter-observer.ti,ab. or intraobserver.ti,ab. or intra- observer.ti,ab. or intertechnician.ti,ab. or

inter-techni- cian.ti,ab. or intratechnician.ti,ab. or intra-technician.ti,ab. or interexaminer.ti,ab. or inter-examiner.ti,ab. or intraex- aminer.ti,ab. or intra-examiner.ti,ab. or interassay.ti,ab. or inter-assay.ti,ab. or intraassay.ti,ab. or intra-assay.ti,ab. or interindividual.ti,ab. or inter-individual.ti,ab. or intraindi- vidual.ti,ab. or intra-individual.ti,ab. or interparticipant.ti,ab. or inter-participant.ti,ab. or intraparticipant.ti,ab. or intra-participant.ti,ab. or kappa.ti,ab. or kappa-s.ti,ab. or kappas.ti,ab. or repeatab*.ti,ab. or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. or generaliza*.ti,ab. or general- isa*.ti,ab. or concordance.ti,ab. or (intraclass and correlation*).ti,ab. or discriminative.ti,ab. or "known group".ti,ab. or factor analysis.ti,ab. or factor analyses.ti,ab. or dimension*.ti,ab. or subscale*.ti,ab. or (multitrait and scaling and (analysis or analyses)).ti,ab. or item discriminant.ti,ab. or interscale correlation*.ti,ab. or error.ti,ab. or errors.ti,ab. or "individual variabil- ity".ti,ab. or (variability and (analysis or val- ues)).ti,ab. or (uncertainty and (measurement or measuring)).ti,ab. or "standard error of measure-ment".ti,ab. or sensitiv*.ti,ab. or responsive*.ti,ab. or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti,ab. or (small* and (real or detectable) and (change or difference)).ti,ab. or meaningful change.ti,ab. or "ceiling effect".ti,ab. or "floor effect".ti,ab. or "Item response model".ti,ab. or IRT.ti,ab. or Rasch.ti,ab. or "Differential item functioning".ti,ab. or DIF.ti,ab. or "computer adaptive testing".ti,ab. or "item bank".ti,ab. or "cross-cultural equivalence".ti,ab. (5772794)

19 (burden of migraine questionnaire or BURMIG or (beck depression inventory or BDI) or (comprehensive headache related quality of life questionnaire or CHQQ) or (completeness of response survey or CORS) or (cognitive impairment scale for migraine attacks or MIG-SCOG) or (chronic pain coping inventory or CPCI) or (depression anxiety stress scale or DASS) or (functional assessment in migraine or FAIM or FAIMQ) or (female sexual function index or FSFI) or (health utilities index or HUI) or (headache impact test or HIT 6 or HIT6 or HIT-6) or (headache impact score or HIS) or (headache management self-efficacy scale or HSES) or (headache-specific locus of control or headache specific locus of control) or (headache disability inventory or HDI) or (headache disability scale or HDS) or ((hospital anxiety and depression scale) or HADS) or ((headache-attributed restriction, disability, social handicap and impaired participation questionnaire) or HARDSHIP or HARDSHIPQ) or (headache activities of daily living index or HADLI) or (headache under response to treatment questionnaire or HURT or HURTQ) or (headache impact questionnaire or HIQ or HImQ) or (headache needs assessment survey or HANA) or (headache intensity or headache duration or headache severity) or (henry ford hospital headache disability inventory or HDI) or (impact of migraine-tension type headache-neck pain or Impact M-TTH-NP) or (italian perceived disability scale or IPDS) or (migraine treatment optimisation questionnaire or M-TOQ-15) or (migraine treatment satisfaction measure or MTSM) or (migraine-specific quality of life scale or MSQoL) or ((migraine work and productivity loss questionnaire) or MWPLQ) or (migraine disability assessment score or MIDAS) or (migraine disability assessment questionnaire or MDAS) or (migraine screen questionnaire or MS-Q) or (migraine impact questionnaire or MIQ) or (migraine specific quality of life questionnaire or MSQL or MSQ or MSQV 2 1) or (24-h MSQoLQ or 24-hour migraine specific quality of life questionnaire) or (patient perception of migraine questionnaire or PPMQ) or (patient health questionnaire or PHQ-9) or (pain disability index or PDI) or (pittsburg sleep quality index or PSQI) or (subjects global impression of change or SGIC) or (pain catastrophizing scale or PCS) or (visual aura rating scale or VARS) or (waters headache questionnaire or WHQ) or (numerical rating scale or NRS or numerical pain intensity scale or numerical pain rating scale or numeric rating scale for pain or NRS pain or NRS-pain) or (visual analogue scale or visual analogue scale for pain or VAS pain or VAS-pain or VAS) or (rating scale or analogue scale) or (SF36 or SF 36 or SF-36 or short form 36 or shortform 36 or short-form 36) or (SF12 or SF 12 or SF-12 or short form 12 or shortform 12 or short-form 12) or (SF6D or SF 6D or SF-6D or short form 6D or shortform 6D or short-form 6D) or (euroqol or euro qol or euro-qol or EQ5D or EQ 5D or EQ-5D)).tw. (315280)

20 (addresses or biography or case reports or comment or directory or editorial or festschrift or interview or lectures or legal cases or legislation or letter or news or newspaper article or patient education handout or popular works or congresses or consensus development conference or consensus development conference, nih or practice guideline).pt. not (*animals/ not *humans/) (3636046)

21 (17 and 18 and 19) not 20 (1566)

22 limit 21 to (english language and humans and "all adult (19 plus years)") (1255)

23 limit 22 to yr="1980 - 2016" (1243)

Appendix 1.3 List of measures included in 'named PROM' searches (EMBASE and MEDLINE) (Total n= 51)

(burden of migraine questionnaire or BURMIG or

(beck depression inventory or BDI) or

(comprehensive headache related quality of life questionnaire or CHQQ) or

(completeness of response survey or CORS) or

(cognitive impairment scale for migraine attacks or MIG-SCOG) or

(chronic pain coping inventory or CPCI) or

(depression anxiety stress scale or DASS) or

(functional assessment in migraine or FAIM or FAIMQ) or

(female sexual function index or FSFI) or

(health utilities index or HUI) or

(headache impact test or HIT 6 or HIT6 or HIT-6) or

(headache impact score or HIS) or

(headache management self-efficacy scale or HSES) or

(headache-specific locus of control or headache specific locus of control) or

(headache disability inventory or HDI) or

(headache disability scale or HDS) or

((hospital anxiety and depression scale) or HADS) or

((headache-attributed restriction, disability, social handicap and impaired participation questionnaire) or

HARDSHIP or HARDSHIPQ) or

(headache activities of daily living index or HADLI) or

(headache under response to treatment questionnaire or HURT or HURTQ) or

(headache impact questionnaire or HIQ or HImQ) or

(headache needs assessment survey or HANA) or

(headache intensity or headache duration or headache severity) or

(henry ford hospital headache disability inventory or HDI) or

(impact of migraine-tension type headache-neck pain or Impact M-TTH-NP) or

(italian perceived disability scale or IPDS) or

(migraine treatment optimisation questionnaire or M-TOQ-15) or

(migraine treatment satisfaction measure or MTSM) or

(migraine-specific quality of life scale or MSQoL) or

((migraine work and productivity loss questionnaire) or MWPLQ) or

(migraine disability assessment score or MIDAS) or

(migraine disability assessment questionnaire or MDAS) or

(migraine screen questionnaire or MS-Q) or

(migraine impact questionnaire or MIQ) or

(migraine specific quality of life questionnaire or MSQL or MSQ or MSQV 2 1) or

(24-h MSQoLQ or 24-hour migraine specific quality of life questionnaire) or

(patient perception of migraine questionnaire or PPMQ) or

(patient health questionnaire or PHQ-9) or

(pain disability index or PDI) or

(pittsburg sleep quality index or PSQI) or

(subjects global impression of change or SGIC) or

(pain catastrophizing scale or PCS) or

(visual aura rating scale or VARS) or

(waters headache questionnaire or WHQ) or

(numerical rating scale or NRS or numerical pain intensity scale or numerical pain rating scale or numeric rating scale for pain or NRS pain or NRS-pain) or

(visual analogue scale or visual analogue scale for pain or VAS pain or VAS-pain or VAS) or

(rating scale or analogue scale) or

(SF36 or SF 36 or SF-36 or short form 36 or shortform 36 or short-form 36) or

(SF12 or SF 12 or SF-12 or short form 12 or shortform 12 or short-form 12) or

(SF6D or SF 6D or SF-6D or short form 6D or shortform 6D or short-form 6D) or

(euroqol or euro qol or euro-qol or EQ5D or EQ 5D or EQ-5D)).tw. (552600)

Appendix 2. Table 1: Multi-item PROMs identified (n=39) from full-text articles assessed for eligibility (searched 1980-2016); n=23 PROMs included in final review ('in-use' 2000-2016).

PROM	**Developer / article in which identified**	**Include / exclude from full review**	**Justification**	**Evaluations included in review (n)**	
Migraine-specific (10)		***5/10***		***1988-1999***	***Post-2000***
BURMIG questionnaire	*Andree, C., M. Vaillant, C. Rott, Z. Katsarava and P. S. Sandor (2008). "Development of a self-reporting questionnaire, BURMIG, to evaluate the burden of migraine." Journal of Headache & Pain 9(5): 309-315.*	*No*	*Developed in Swiss population – evidence of translation into English not clear in article. Initial evaluation in Swiss population. Exclude.*	-	*(1)*
Functional Assessment in Migraine Questionnaire (FAIM)	Pathak, D. S., D. J. Chisolm and K. A. Weis (2005). "Functional Assessment in Migraine (FAIM) questionnaire: development of an instrument based upon the WHO's International Classification of Functioning, Disability, and Health." Value in Health **8**(5): 591-600.	Yes	Fulfils inclusion criteria	-	1
Headache Needs Assessment (HANA) survey	Cramer, J. A., S. D. Silberstein and P. Winner (2001). "Development and validation of the headache needs assessment (HANA) survey." Headache **41**(4): 402-409.	Yes	Fulfils inclusion criteria	-	1
Migraine Disability Assessment Scale (MIDAS)	Stewart, W. F., R. Lipton, K. Kolodner, J. Liberman and J. Sawyer (1999A). "Reliability of the migraine disability assessment score in a population- based sample of headache sufferers." Cephalalgia **19**(2): 107-114.	Yes	Fulfils inclusion criteria	2	10
Migraine Quality of life Questionnaire (MQoLQ)	*Hartmaier, S. L., N. C. Santanello, R. S. Epstein and S. D. Silberstein (1995). "Development of a brief 24-hour migraine-specific quality of life questionnaire." Headache 35(6): 320-329.*	*No*	*No evaluations identified post-2000. Used as a comparator in establishing evidence in support of a new measure – the MTSM – but this is limited [32].*	*(1)*	*0*

Migraine-specific Quality of Life Questionnaire (MSQ) v1.	*Jhingran, P., J. T. Osterhaus, D. W. Miller, J. T. Lee and L. Kirchdoerfer (1998). "Development and validation of the migraine-specific quality of life questionnaire." Headache **38**(4): 295-302.*	*No*	*Succeeded by MSQ v2.1*	*(1)*	-
Migraine-specific Quality of Life Questionnaire (MSQ) v2.	*Jhingran, P., S. M. Davis, L. M. LaVange, D. W. Miller and R. W. Helms (1998). "MSQ: Migraine-specific quality-of-life questionnaire: Further investigation of the factor structure." PharmacoEconomics **13**(6): 707-717.*	*No*	*Succeeded by MSQ v2.1*	*(1)*	-
Migraine-specific Quality of Life Questionnaire (MSQ) v2.1.	Martin, B. C., D. S. Pathak, M. I. Sharfman, J. U. Adelman, F. Taylor, W. J. Kwong and P. Jhingran (2000). "Validity and reliability of the migraine-specific quality of life questionnaire (MSQ Version 2.1)." Headache **40**(3): 204-215.	Yes	Fulfils inclusion criteria	-	9
Migraine-specific Quality of Life Questionnaire (MSQoL)	Wagner, T. H., D. l. Patrick, B. S. Galer and R. A. Berzon (1996). "- A new instrument to assess the long-term quality of life effects from migraine: development and psychometric testing of the MSQOL." Headache **36**(8): 484-492.	Yes	Fulfils inclusion criteria	2	1
Migraine Symptom Frequency Bother questionnaire	*Patrick, D. L., M. L. Martin, D. M. Bushnell and J. Pesa (2003). "Measuring satisfaction with migraine treatment: expectations, importance, outcomes, and global ratings." Clinical Therapeutics **25**(11): 2920-2935.*	*No*	*Ad hoc measure developed specifically for single study. Not evaluated or applied again.*	*-*	*1*
Headache-specific (8)		**6/8**			
Comprehensive Headache Related Quality of Life Questionnaire (CHQQ)	*Manhalter, N., G. Bozsik, A. Palasti, E. Csepany and C. Ertsey (2012). "The validation of a new comprehensive headache-specific quality of life questionnaire." Cephalalgia **32**(9): 668-682.*	*No*	*Non-Anglicised evaluations (n=2)*	*-*	*(2)*

EUROLIGHT questionnaire	Andree, C., Vallaint M., Barre, J.,Katsarava R. et al. (2009). "Development and Validation of the EUROLOGHT questionnaire to evaluate the burden of primary headache disorders in Europe." Cephalalgia 30 (9):1082-1100.	Yes	Fulfils inclusion criteria	-	1
Headache Activities of Daily Living Index (HADLI)	Vernon, H. and G. Lawson (2015). "Development of the headache activities of daily living index: Initial validity study." Journal of Manipulative and Physiological Therapeutics **38**(2): 102-111.	Yes	Fulfils inclusion criteria	-	1
Headache Attributed Restriction, Disability, Social Handicap and Impaired *Participation (HARDSHIP) questionnaire*	*Steiner, T. J., G. Gururaj, C. Andree, Z. Katsarava, I. Ayzenberg, S. Y. Yu, M. Al Jumah, R. Tekle-Haimanot, G. L. Birbeck, A. Herekar, M. Linde, E. Mbewe, K. Manandhar, A. Risal, R. Jensen, L. P. Queiroz, A. I. Scher, S. J. Wang and L. J. Stovner (2014). "Diagnosis, prevalence estimation and burden measurement in population surveys of headache: presenting the HARDSHIP questionnaire." Journal of Headache and Pain **15**(1).*	*No*	*Interview-administration only – modular instrument: demographic, diagnostic, headache-attributed burden – symptoms, health-care utilisation, disability, productive time loss, impact on education, career and earnings, control, relationships and family, qol, well-being, co-morbidities.*	-	-
Headache Disability Impact Questionnaire (HDI)	Niere K, Quin A. Development of a headache-specific disability questionnaire for patients attending physiotherapy. *Man Ther.* 2009 Feb;14(1):45-51	Yes	Fulfils inclusion criteria	-	1
Headache Impact Test (HIT)	Bjorner, J. B., M. Kosinski and J. E. Ware Jr (2003A). "Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the Headache Impact Test (HITTM)." Quality of Life Research **12**(8): 913-933.	Yes	Fulfils inclusion criteria	-	3
Headache Impact Test-6 (HIT-6)	Bjorner, J. B., M. Kosinski and J. E. Ware Jr (2003C). "Using item response theory to calibrate the Headache Impact Test (HITTM) to the metric of traditional headache scales." Quality of Life Research **12**(8): 981-1002.	Yes	Fulfils inclusion criteria	-	12
Henry Ford hospital headache disability inventory (HDI).	*Jacobson, G. P., N. M. Ramadan, S. K. Aggarwal and C. W. Newman (1994). "The Henry Ford hospital headache disability inventory (HDI)." Neurology **44**(5): 837-842.*	*No*	*No evaluations identified post-2000*		-

SF-36 'Headache-specific' Modification	Magnusson JE, Riess CM, Becker WJ. Modification of the SF-36 for a headache population changes patient-reported health status. *Headache.* 2012; 52(6): 993-1004.	Yes	Fulfils inclusion criteria	-	1
Response to treatment (7)		**6/7**			
Completeness of Response Survey (CORS)	Coon, C. D., S. E. Fehnel, K. H. Davis, M. C. Runken, M. E. Beach and R. K. Cady (2012). "The development of a survey to measure completeness of response to migraine therapy." Headache **52**(4): 550-572.	Yes	Fulfils inclusion criteria	-	1
Migraine Assessment of Current Therapy (Migraine-ACT) Questionnaire	Dowson, A. J., S. J. Tepper, V. Baos, F. Baudet, D. D'Amico and S. Kilminster (2004). "Identifying patients who require a change in their current acute migraine treatment: The Migraine Assessment of Current Therapy (Migraine-ACT) questionnaire." Current Medical Research and Opinion **20**(7): 1125-1135.	Yes	Fulfils inclusion criteria	-	2
Migraine Therapy Assessment Questionnaire (M-TAQ)	Chatterton ML[1], Lofland JH, Shechter A, Curtice WS, Hu XH, Lenow J, Smullens SN, Nash DB, Silberstein SD. Reliability and validity of the migraine therapy assessment questionnaire. Headache. 2002 Nov-Dec;42(10):1006-15.	Yes	Fulfils inclusion criteria	-	1
Migraine Therapy Optimisation Questionnaire (M-TOQ)	Lipton, R. B., K. Kolodner, M. E. Bigal, D. Valade, M. J. A. Lainez, J. Pascual, A. Gendolla, G. Bussone, N. Islam, K. Albert and B. Parsons (2009). "Validity and reliability of the migraine-treatment optimization questionnaire." Cephalalgia **29**(7): 751-759.	Yes	Fulfils inclusion criteria	-	1
Migraine Treatment Satisfaction Measure (MTSM)	Patrick, D. L., M. L. Martin, D. M. Bushnell and J. Pesa (2003). "Measuring satisfaction with migraine treatment: expectations, importance, outcomes, and global ratings." Clinical Therapeutics **25**(11): 2920-2935.	Yes	Fulfils inclusion criteria	-	2

Patient Perception of Migraine Questionnaire (PPMQ)	*Davis, K. H., L. Black and B. Sleath (2002). "Validation of the Patient Perception of Migraine Questionnaire." Value in Health **5**(5): 422-430.*	*(Yes)*	*Fulfils inclusion criteria – but succeeded by PPMQ-R*	-	*1*
Patient Perception of Migraine Questionnaire – Revised (PPMQ-R)	Revicki, D. A., M. Kimel, K. Beusterien, J. W. Kwong, J. A. Varner, (2006).	Yes	Fulfils inclusion criteria	-	2
Generic (8)		**6/8**			
Generic quality of life / health status					
Profile measures (4)		**3/4**			
Short Form Health Survey 8 (SF-8)	Turner-Bowker, D. M., M. S. Bayliss, J. E. Ware Jr and M. Kosinski (2003). "Usefulness of the SF-8TM Health Survey for comparing the impact of migraine and other conditions." Quality of Life Research **12**(8): 1003-1012.	Yes	Fulfils inclusion criteria	-	4
Short Form Health Survey 12 (SF-12)	Lipton, R. B., S. W. Hamelsky, K. B. Kolodner, T. J. Steiner and W. F. Stewart (2000). "Migraine, quality of life, and depression: a population-based case-control study." Neurology **55**(5): 629-635.	Yes	Fulfils inclusion criteria	-	1
Short Form Health Survey 36 (SF-36)	Solomon, G. D., F. G. Skobieranda and L. A. Gragg (1993). "Quality of life and well-being of headache patients: measurement by the medical outcomes study instrument." Headache **33**(7): 351-358.	Yes	Fulfils inclusion criteria	-	5
World Health Organisation Disability Assessment II (WHO-DAS II)	*Raggi, A., M. Leonardi, G. Bussone and D. D'Amico (2011). "Value and utility of disease-specific and generic instruments for assessing disability in patients with migraine, and their relationships with health-related quality of life." Neurological Sciences 32(3): 387-392.*	*No*	*Non-Anglicised evaluations (n=3)*	-	*3*

Utility measures (4)		**3/4**			
EuroQol EQ-5D-3L	Essink-Bot, M. L., P. F. Krabbe, G. J. Bonsel and N. K. Aaronson (1997). "An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument." Medical care **35**(5): 522-537. *Non-Anglicised evaluation* *Anglicised evaluations included in review (n=3)*	Yes	Fulfils inclusion criteria	-	3
Health Utilities Index (HUI)	*Mo, F., B. C. Choi, F. C. Li and J. Merrick (2004). "Using Health Utility Index (HUI) for measuring the impact on health-related quality of Life (HRQL) among individuals with chronic diseases." The Scientific World Journal **4**: 746-757.*	*No*	*Succeeded by HUI-3*	-	*1*
Health Utilities Index-3 (HUI-3)	Brown, J. S., P. J. Neumann, G. Papadopoulos, G. Ruoff, M. Diamond and J. Menzin (2008). "Migraine frequency and health utilities: findings from a multisite survey." Value in Health 11(2): 315-321.	Yes	Fulfils inclusion criteria	-	(1)
Quality of Well Being Scale (QWB-8)	Sieber, W. J., K. M. David, J. E. Adams, R. M. Kaplan and T. G. Ganiats (2000). "Assessing the impact of migraine on health-related quality of life: An additional use of the quality of well-being scale-self-administered." Headache **40**(8): 662-671.	Yes	Fulfils inclusion criteria	-	1

Appendix 3: Table 2: Assessment criteria for the quality of reported measurement properties [17,19,20].

Measurement property	Rating	Assessment of quality
Reliability		
Internal consistency	+	Cronbach's alpha(s) > 0.70
- *the extent to which items within a measure are internally consistent*	?	Cronbach's alpha not evaluated or dimensionality unknown
	-	Cronbach's alpha(s) < 0.70
Reliability	+	Intra-class Correlation Coefficient (ICC)/weighted Kappa >0.70 OR Pearson's r >0.80
(test-retest / inter-rater / inter-rater)	?	Neither ICC/weighted Kappa, not Pearson's r evaluated
- *the extent to which a measure provides the same results on repeated completions, assuming no change in the underlying health state*	-	ICC/weighted Kappa <0.70 OR Pearson's r <0.80
Validity		
Content validity	+	Authors provide a clear description of the measurement aim, target population, concept(s) measured and process of item selection. Members of the target population and experts in the field were clearly identified as being involved in development. For measures applied for the first time in a new population, evidence that the views of members of the target population (and experts in the field) have been sought to determine relevance, comprehension and comprehensiveness.
- *the extent to which the item content of a measure is an adequate reflection of the construct being measured*	?	Insufficient evidence available
	-	No detail re measurement aim, target population, concept(s) measured, process of item selection; members of the target population or experts were not specifically involved in development. For measures applied for the first time in a new population, evidence whereby the relevance and acceptability of the measure with members of the target audience or experts was not provided.
Construct validity - Structural validity	+	Factors should explain 50% of the variance
- *the extent to which PROM scores adequately reflect the dimensionality of the construct being measured.*	?	Explained variance not reported
	-	Factors explain < 50% of the variance
Construct validity - Hypothesis testing	+	Correlations with measures of the same construct should be >0.50 OR at least 75% of the results in accordance with hypothesized associations AND correlations with related constructs should be higher than with those reported with unrelated constructs
- *convergent (the extent to which measures of related constructs are related to each other)*	?	Only report correlations with unrelated constructs

- *discriminant (the extent to which a measure can demonstrate differences between groups known to differ on important variables)*	-	Correlations with measures of the same construct are <0.50 OR < 75% of the results in accordance with hypothesized associations OR correlations with related constructs are lower than those reported with unrelated constructs
Responsiveness - *the ability to detect important change over time in the construct being measured (criterion / construct-based assessment)*	+	Change-score correlations with measures of the same construct are >0.50 OR at least 75% of the results are in accordance with hypothesized associations OR the Area Under the Curve (AUC) is >0.70 AND change-score correlations with measures of related constructs are higher than those reported with unrelated constructs
	?	Solely correlations with unrelated constructs
	-	Change-score correlations with measure of the same construct <0.50 OR < 75% of the results are in accordance with hypothesized associations OR AUC is <0.70 AND change-score correlations with related constructs are lower than those reported with unrelated constructs

Appendix 4: Table 3 Characteristics of reviewed PROMs evaluated in the headache population (total = 23)

PROM ***(Author; web-link[b]; completion format)*** [c]	**Items n**[a]	**Construct** **Domains (items)**	**Response options (range)**	**Recall Period**	**Score range**	**Administration**
Condition-specific (17)						*Time*
Migraine-impact (5/17)						
Functional Assessment in Migraine (FAIM) (Pathak et al, 2005)[33] *Self-completion* *Items listed in development paper*	9 + 5	Underpinned by the WHO ICF. Focus on the functional impact of migraine. 3 domains: Two mental function: Attention/Thought (5): concentration, control of life, focus on issues, spontaneity, think quickly. Perception (4): find a peaceful place, light/sound/interaction avoidance. One overall domain: 'Activity and Participation'(5): select up to 5 items from list of 28.	Item stem: How much does each item impact on their lives? 7-point scale: 1= 'not at all' to 7 – 'all of the time'	Within 24-hours of their typical migraine onset	Item summation within the three domains. A/T: range 5-35 P: range 4-28 A/P: range 5-35 Lower scores indicate less functional impact.	5-10 minutes *Not reported in headache population*
Headache Needs Assessment (HANA) (Cramer et al,2001)[34] *Self-completion* *Copy of PROM in appendix to paper*	7	Migraine quality of life – frequency and bothersomeness (7): Anxiety/worry; depression/discouragement; self-control; energy; function/work; family/social activities; overall impact of migraines.	For each item: Frequency: How often has this problem occurred?: never / rarely / sometimes / often / all the time) Bothersomeness: How much has this problem bothered you?: not at all / a little / some / a lot / a great deal.	Not stated	Item summation. Total range 7-175, where lower scores indicate less impact.	5-10 minutes *Not reported in headache population*
Migraine Disability Assessment Score (MIDAS) (Stewart et al. 1999)[35] http://www.achenet.org/midas/	12	Migraine disability (*but attribution is 'headache')* 3 domains (5 scored items) Missed days/ reduced productivity at paid work (2)	Item stem: About ALL of the headaches you have had… Frequency - number of days/ half days of disability	3 months	Total score derived as sum of lost days, where greater number of days indicates greater migraine-related disability.	5-10 minutes *Not reported in headache population*

Self-completion		Missed days/ reduced productivity at household work (2) Missed non-work activities (1) Plus: 2 unscored items Frequency of headaches *(how many days?)* (1) Headache pain severity (1)	 Not scored: Number of days Scale 0-10 (0= no pain at all, and 10= pain as bad as can be)			
Migraine-Specific Quality of Life Questionnaire version 2.1 (MSQ v2.1) (Martin et al, 2000 [36] http://www.outcomes-trust.org/instruments.htm#msql2.1 *Contact for further details and copy of questionnaire:* michael.c.runken@gsk.com *Self-completion*	9	Health-related quality of life - impairments attributed to migraine 3 domains (14 items) Role function – Restrictive (RR)(7): social – family/friends; leisure (2); work/ADL(3); cognition (1); symptoms: fatigue (1) Role function – Preventive (RP)(4): symptoms: fatigue (1); work /ADL (2); social (1) Emotional function (EF)(3): frustration (1); feeling a burden (1); letting others down (1)	 Item stem: How often have migraines interfered with / limit/ed your ability to / keep you from getting as much done / had difficulty in / How often have you had to cancel / need help / have to stop / not able to go… … because of your migraine 6-point categorical scale: None of the time (1), A little bit of the time (2), Some of the time (3), A good bit of the time (4), Most of the time (5), All of the time (6).	4 weeks	Items summed within the three domains and transformed to 0-100 scale. Higher scores indicate a worse quality of life	5-10 minutes *Not reported in headache population*
Migraine-Specific Quality of life (MSQOL) measure (McKenna et al, 1998) [37]	1	Needs-based migraine-specific quality of life 3 domains (20 items):	Item stems: Various – include: 'I try to avoid …' / 'It's important for	Responders advised to 'choose the answer that	All items summed (score range 20-80) and transformed to 0-100 scale.	5-10 minutes

http://www.galen-research.com/content/measures/MSQoL%20UK%20-%20First%20page%20sample.pdf *Payment required to access full version of the questionnaire* *Self-completion*		Avoidance behaviours (10) Social relations (6) Feelings (4)	me…' / 'I feel helpless …'/ 'I worry about …' 4-point categorical response scale: range: Yes, very much / I try very hard / very important (1) Yes, quite a lot / I try quite hard / quite important (2) A little / I do not try very hard / not very important (3) Not at all / I do not try at all / it's not important at all (4)	applies to you: between migraine attacks OR with any treatment you use now.	Higher scores indicate better quality of life	*Not reported in headache population*
Headache-impact (6/17)						
EUROLIGHT (Andree et al, 2010)[28] http://www.l-t-b.org/index.cfm/spKey/publications.html *Self-completion*	1	Burden of primary headache disorders. Includes assessment of headache characteristics, co morbidities, disease management and quality of life 6 sections (103 items): 1) Biographical (age, gender, language and employment) 2) Screening questions for headache (life-time and 1-year prevalence) 3) Diagnostic questions - based on the criteria of the *International Classification of Headache Disorders (*ICHD-II) 4) Questions about any headache experienced 'yesterday' (point prevalence) 5) Use of healthcare resources (medicines, investigations, consultations, etc.)	All domains categorical response categories – various number of options	Headache frequency in past month/yesterday Healthcare past 30 days Headache impact on 'own life' Headache-related lost time in past 3 months	Various As per WHOQoL, HALT-index and HADS.	Not reported. Will require considerable completion time

		6) Impact of headache on work, family life and social activities: includes items taken from the WHOQOL (8-items), the HALT-index and HADS.				
Headache Activities of Daily Living Index (HADLI) (Vernon & Lawson, 2015) [29] *Self-completion*	1	Headache-related 'activity disability' - ability of an individual to engage with usual activities of daily life during headache episode 1 domain (9 items): Personal care Lifting Reading (including computers) Sleeping (over last week) Exercising (over last week) Social activities Work Driving or travelling Recreation	6-point categorical response options, where 0 is best ability and 5 is worst ability.	During headache episode 'when you have a headache'	Item summation: score range 0 to 45, where 45 is maximum activity disability. Total score converted to percentage	3 minutes
Headache Disability Questionnaire (HDQ) (Niere & Quin, 2009) [30] *Self-completion*	1	Headache specific disability in patients presenting for physiotherapy treatment 3 domains (9 items): Pain (2): usual pain intensity; when pain is severe. Activity Limitation (4): Decreased efficiency in non-work activities Decreased ability to work/study Decreased efficiency in housework or chores Proportion of times when work is missed Activity Prevention (3): Number of days where chores prevented	11 point Numerical Rating Scales Anchors (0-10): Adjectival anchors: No pain (0) - Worst pain (10) Never (0) – Always (10) None (0) – Everyday (10) Not reduced (0) - Unable to work (10) Not reduced (0) - Unable to perform (10)	Past 1-month	Item summation. Index range 0-90, where higher scores indicate greater headache-specific disability.	??

		Number of days non-work activities prevented Number of days in last month when had to lie down for >1 hour;				
Headache Impact Test (HIT) CAT-HIT *(IRT-HIT)* (Bjorner JB et al. 2003a [3]; Ware JE et al, 2003 [57]) *Self-completion- requires internet interface for CAT completion*	3	Headache impact Items, derived from four established measures: MIDAS MSQ (v1.0) Headache Disability Index (HDI) Headache Impact Questionnaire (HIMQ) Plus experimental items generated from clinical trial data, and consultation with clinicians. One domain 'Headache Impact': 54-item 'item bank' Items cover a wide spectrum of headache impact, including minor headache. Items cover of pain, role and social functioning, energy/fatigue, cognitive function, and mental health	Up to five categorical responses Internet completion only using Computerized Adaptive Testing (CAT): CAT-HIT	1-month	Scored using Item Response Methods (IRT) *(also referred to as IRT-HIT).* Number of completed items determined by 'stopping rule': mean number 6/54 items. Norm-based scoring with mean 50 (SD 10), where higher scores indicate very severe headache impact.	Approximate response times of 1.5 minutes for those with least headache impact (HIT scores < 50) Range from 2.4 items/ minute for a 9-item survey to 3.3 items/minute for a 6-item survey
Headache Impact Test (HIT-6) (Kosinski et al, 2003a) [31] https://www.optum.com/optum-outcomes/what-we-do/disease-specific-health-surveys/hit-6.html *Self-completion*	8	Headache impact Static, short-form HIT: 6 domains (6-items) Pain (*headache - how often is the pain severe*?) (1) Social functioning (*limit your ability to do usual daily activities – work / adl / social?)* (1)	Equally weighted 5-option categorical scale with specific item score (generated to closely match the IRT score) Item stem: *'how often…?'* Attribution: *'when you have a headache /*	3 items (Vitality; Psychological distress; Cognition): past 4weeks 2 items (Pain. Role limitation): 'when you have a headache'	Item summation to create index score: range 36-78 Score interpretation (norm-based mean 50 (SD 10)): >60: very severe impact 56-59: substantial impact	

		Role functioning *(how often do you wish you could lie down?)* (1) Vitality *(too tired to do work or adl?)* (1) Cognitive functioning *(limit ability to concentrate on work or adl?)* (1) Psychological distress *(felt fed up or irritated)* (1)	*because of your headache'* Range *(weighted item responses):* Never (6 points), Rarely (8 points), Sometimes (10 points), Very Often (11 points), Always (13 points)	1 item (Social function): how often do headaches limit …	50-55: some impact 49 or <: little or no impact	
'Headache' SF-36 (Magnusson J.E. et al. 2012)[32] *Self-completion*	1	Headache-related health status Modification of original SF-36 to 'improve applicability' to the headache population by inserting 'including your headaches' to 6/36 items: physical functioning (item 3); role limitation - physical (item 4); social functioning (items 6 and 10); bodily pain (items 7 and 8)	Equally weighted 3 or 5 point Likert Scale	4 weeks or 1 week	Scale scores transformed to 0-100 calibrated at 50 as the norm 8 domains 2 summary scales	5-10 minutes
Response to migraine-specific treatment (6/17)						
Completeness of Response to Migraine Therapy Survey (CORS) (Coon CD. et al. 2012) [38] *Both versions of CORS illustrated in full in appendix to article* *Self-completion*	1	Optimal treatment: considers factors important to patients when considering the initiation/continuation of migraine treatment Two modules: 1.'Static CORS' – to evaluate treatment at a single time-point 5 domains (24 items): Frequency of Complete Relief (FCR) (6 – *how often complete*	*Item stem and response options:* 1.'Static CORS': categorical response options range 3, 4 or 5-options: FCR: *how often does your current M Rx*	'overall experience with your current migraine treatment'	1.'Static CORS': three domains scores based on item summation FCR (6): 1a,2b,3b,4b,5b,6b SCR (6): 1b,2c,3c,4c,5c,6c SRF (4): 7,8,9,10 - where higher scores indicate better medication response 2.'Comparative CORS': item summation to	10-15 minutes *Not reported in headache population*

		relief of symptoms* plus irritability /moodiness**) Speed of Complete Relief (SCR) (6 - *how quickly complete relief of* symptoms* plus irritability/ moodiness**) Speed of Return to Functionality (SRF)(4 – able to concentrate/ think; normal activities; functioning normally (100%); feeling completely normal (100%)) Frequency of Migraine Recurrence (FMR)(1) Confidence in Treatment (CIT)(2) 5 additional items describe presence of symptoms*: 24-items address: Symptoms* - Headache-specific pain; neck/shoulder pain; nausea; sensitivity to light; sensitivity to sound (5 items). Emotional well-being – **irritability or moodiness Cognition – ability to 'concentrate or think' ADL – resumption of normal activities Function – resumption of normal functioning (100%) Feeling 'completely normal' (100%) Confidence in current medical: frequency of return of M	*completely relieve your X?:* 5-point (0-4): range 0= none of the time, to 4 = all or almost all of the time. SCR: *how quickly does your current M Rx completely relieve your X?:* 4-point (1-4): range 4 = < 30 mins, to 1 = > 2hrs SRF: *how quickly are you able to … after taking your current M RX?:* 5-point (1-5): range 5 = < 30 mins, to 1 = > 4hrs FMR: 5-point (0-4): range 0= none of the time, to 4 = all or almost all of the time. CIT: 3-point (0-2): range 0 = not at all confidence, 1 = somewhat confident; to 2 = very confident.		produce index score (range 8 to 40), where higher scores suggest a better response to current medication than previous	

		within 24hrs; confidence that Rx will completely relieve M; confidence that M will not come back. 2. 'Comparative CORS (8 items) – for the comparative evaluation of two treatments at a single time-point. 8 domains (8 items): Completeness of Relief (1) Speed of Relief (1) Persistence of Relief (1) (e.g. prevented symptoms from coming back within 24-hrs) Return to Normal Function (1) Fatigue (1) Confidence in Treatment (2): that one does would completely relieve M within 2-hrs; that M would not come back within 24-hrs Overall Satisfaction (1) – most satisfied	2. 'Comparative CORS': *Which medication provided the complete / quicker / longer-lasting / allowed more normal function / experience less fatigue / feel more confidence that one does would completely relieve your M within 2 hours?:* 5-point categorical scale (1-5); where a score of '1' favours the previous medication; a score of 3 suggests no preference between medications; a score of 5 favours the study medication.			
Migraine – Assessment of Current Therapy (M-ACT) *Copy included in publication* (Dowson et al, 2004)[39] *Self-completion*	2	4 domains (27 long-form/ 4 short): Headache impact (11/1) Global assessment of relief (9/1) Consistency of response (3/1) Emotional response (4/1)	Item stem: 'When you take your treatment: Dichotomous answers: Yes/No, where Yes = 1 and No = 0	Varies: between 2hours and 48-hours	Item summation	5-10 minutes *Not reported in headache population*
Migraine Therapy Assessment Questionnaire (M-TAQ)	1	To identify barriers to optimal migraine management and improve patient outcomes	Dichotomous answers: Yes/No,	Varies – mostly 2 hours.	Item summation Range 0 to 8 (items 3 and 4 scored together),	5-10 minutes

(Chatterton et al, 2002)[40] *Self-completion* *Copy of PROM included in appendix to paper*		A disease-management/screening tool to identify individuals whose migraine management is sub-optimal (9) Migraine control, frequency of attacks, knowledge and behavioural barriers, economic burden, treatment satisfaction	where Yes = 1 and No = 0		where higher scores indicates greater number of migraine issues Also 3 domains: Migraine control; Knowledge/behaviour /treatment satisfaction; Economic burden,	*Not reported in headache population*
Migraine-Treatment Optimization Questionnaire (M-TOQ) (Lipton RB. et al. 2009) [41] *Self-completion*	1	Aims to support treatment optimization – defined as the achievement of realistic treatment goals. The M-TOQ was developed to provide a rapid assessment of migraine therapy for use in primary care settings M-TOQ 15: 5 domains (15 items): Functioning (3) Rapid relief of headache (3) Consistency of response (3) Prevention of recurrence (3) Side effects (3) M-TOQ 5: 5 domains (5 items) Functioning (1) Rapid relief of headache (1) Consistency of response (1) Prevention of recurrence (1) Side effects (1)	Dichotomous answers: Yes/No, where Yes = 1 and No = 0	Last 4 weeks	M-TOQ 15 Item summation producing five domain scores or an index score. High scores suggest good response to treatment, suggesting that treatment change is *unlikely* to be required. M-TOQ 5 Item summation producing an index score. If answer 'Yes' to all five items - treatment is considered satisfactory. If answer 'no' to any single question, a change in treatment should be considered. *A 'treatment optimization' table is provided to support score interpretation and clinical decision-making (Table 5)[24]*	5-10 minutes *Not reported in headache population*
Migraine Treatment Satisfaction Measure (MTSM)	2	Migraine treatment satisfaction	TE-M: 5-point response scale (1-5) where 1 is the worst	At onset and 24 hours after migraine episode	Overall 'MTSM' treatment satisfaction score is generated as	15-20 minutes

(Patrick et al, 2003) [42] *Self-completion*		4-part assessment: 1) Expectations of Treatment for Migraine (TE-M); 9 items) – worded to express the 9 attributes as an 'expectation'. 2) Importance Ranking for Migraine Treatment (IR-M); 9 items) – respondents rank items to express their 'desired expectations; for each attribute. 3) Outcomes of Treatment for Migraine (TO-M); 9-items – correspond to the TE-M items) – produces a self-report of treatment outcome for each attribute. 4) Satisfaction with Migraine Treatment (PST-M; 9-items) – reflect satisfaction with treatment outcome across the 9 attributes. 9 attributes ('items') associated with migraine relief: Pain relief Speed of relief Freedom from pain Additional symptoms Confidence in treatment Disruption in life Dosing Freedom from relapse Ease of use	case scenario (eg, no relief) and 5 is the best (eg total relief). IR-M: ranking items on a 10cm line (where 0 = not important and 10 = most important). Intersection with the line = score (range 0-10). TO-M: 5-point response scale (1-5) reflecting actual outcome, where 1 =- worst case scenario (eg no relief) to 5 = best case (eg total relief). PST=M: 10cm VAS where 0 = most dissatisfied and 10 = most satisfied.		the sum of the nine derived attribute scores (a detailed scoring procedure is detailed by the developers [27]): score represents patients expectations about Rx, modified by Rx experience, weighted by their adjusted importance values, and used to modify the raw satisfaction values [27]. Scores also calculable for three domains: Expectations (TE-M) Outcomes (TO-M) Satisfaction (PST-M)	*Not reported in headache population*
Patient Perception of Migraine Questionnaire -Revised (PPMQ-R) (Revicki DA et al. 2006)[43]	2	Patient satisfaction with acute migraine therapy Core: 4 domains (19 items) Efficacy: satisfaction with treatment efficacy (11) Function: ability to perform usual activities (4) Ease of Use (2)	Core domains 1 to 4 and 3 global items: 7-point 'Likert' scale: range 1 = very satisfied to 7 = very dissatisfied. Domain 5 – 'Bothersomeness': 5-	4 weeks	Item summation to create domain scores: Domain 1: range 11 to 77 Domain 2: range 4 to 28 Domain 3: range 2 to 8 Domain 4: range 2 to 8: where lower scores	

		Cost (2) (*may be removed where 'cost' is not a consideration*). Additional fifth domain (10 items): 'Bothersomeness' or 'Tolerability' - related to side effects (10) 3 global items: Overall satisfaction with medication effectiveness Side effects General treatment	point 'Likert' scale: range 1 = not at all, to 5 = extremely.		suggest greater satisfaction with treatment. Item summation of three core domains (Efficacy, Function and Ease) to produce 'Total Satisfaction Score': score transformed to 0-100, where higher scores represent greater satisfaction.	
Generic measures (6)						
Profile measures (3/6)						
Short Form 36-item Health Survey (SF-36) (version 1 (v1)) [Ware et al, 1994] [44] https://campaign.optum.com/content/optum/en/optum-outcomes/what-we-do/health-surveys.html *Self-completion or interview administered*		General health status 8 domains (36 items) Bodily pain (BP)(2) General health (GH)(5) Mental health (MH) (5) Physical functioning (PF)(10) Role limitation-emotional (RE)(3) Role limitation-physical (RP)(4) Social functioning (SF)(2) Vitality (V)(4)	Categorical: 2-6 options	Recall: Standard 4-weeks Acute 1-week	Requires algorithm to score domains Norm-based scoring: score transformed to 0-100 (mean 50 (SD 10)) Individual domain scores ('profile') or 2 summary scales: Physical Component Summary Mental Component Summary	15 to 30 mins *Not reported in headache population*
Short Form 12-item Health Survey (SF-12)(v1) [Ware et al, 1995][45] https://campaign.optum.com/content/optum/en/optum-outcomes/what-we-do/health-surveys.html	1	Health Status 8 domains (12 items) Physical functioning ((n items per domains???) Social functioning Role physical Bodily pain	Categorical: 2-6 options	Recall: Standard 4-weeks Acute 1-week	Requires algorithm to score domains Norm-based scoring: score transformed to 0-100 (mean 50 (SD 10)) 2 summary scales:	5-15 minutes *Not reported in headache population*

Self-completion or interview administered		Mental health Role emotional Vitality General health			Physical Component Summary Mental Component Summary	
Short Form 8-item Health Survey (SF-8)(v1) [Ware et al, 2001][46] https://campaign.optum.com/optum-outcomes/what-we-do/health-surveys.html?gclid=CPj1nb6YoM8CFXQo0wodZXEDLQ	1	Health Status 8 domains (8 items) Physical functioning (1) Social functioning (1) Role physical (1) Bodily pain (1) Mental health (1) Role emotional (1) Vitality (1) General health (1)	Categorical: 2-6 options	Recall: Standard 4-weeks Acute 1-week	Requires algorithm to score domains Norm-based scoring: score transformed to 0-100 (mean 50 (SD 10)) 2 summary scales: Physical Component Summary Mental Component Summary	5-10 minutes *Not reported in headache population*
Utility measures (3/6)						
EuroQoL EQ-5D (3L) (EuroQoL Group, 1990)[47] http://www.euroqol.org/ *Self-completion or interview administered*	3	Quality of Life 5 domains (5 items) Mobility Self-care Usual activities Pain/discomfort Anxiety/depression	3-point descriptive response options: no problems, some problems, severe problems.	Today	Utility index value (society assigned value system algorithm): -0.59 to 1.00 where 1.00 is perfect quality of life, 0 is death, and <0 is a health state worse than death	2 to 5 mins *Not reported in headache population*
Health Utility Index – 3 (HUI-3) (Feeney et al, 2002)[48] http://www.healthutilities.com/hui3.htm *Self-completion or interview administered*	1	Multi-attribute health status classification system Describes the comprehensive health state of an individual as 8 domains (attributes) (8 items): Vision Hearing Speech Ambulation Dexterity Emotion Cognition	1 to 5 or 1 to 6 descriptive response options per attribute / domain, where 1 is best health, 5 or 6 is worst health.	Current	Standard algorithms. 0 to 1.00 where 1.00 is perfect QoL	5 mins *Not reported in headache population*

		Pain				
Quality of Well-being Scale (QWB) (QWB - Kaplan et al, 1993 [49]; QWB-SA - Andresen et al, 1998 [50]) Interview and self-administered (SA)	1	Generic measure of HRQoL; used to calculate QALYs QWB – interview administered *(fewer than QWB-SA)* QWB-SA – self-administered (77 items) Symptom scale 3 scales of function: mobility, physical activity, social activity		6-days 3-days	Overall score based on a preference-weighted average functioning in previous 6-days. Utility index score: 0.0 (death) to 1.0 (perfect health)	QWB-SA: average 11 minutes

Appendix 5: Table 4: Characteristics of included studies (n= 46)

Study (Author;yr) [ref]	Country[g] Language[g]	Population and Headache definition	Study: design, setting, sample size[g]	Mean [g]age (SD; range)	Gender distribution (F Female)[g]	Treatment descriptions[g]	% missing responses: acceptable?[g *]	PROM focus; Additional info [g]
Andree et al 2010 [28]	Five countries: UK, Italy, Spain, Germany/Austria, France	Patients from mixed settings with diagnosis of headache: International Classification of Headache Disorders (ICHD-II) 2004	International cross-sectional survey: Population recruitment: UK and Italy: Headache/migraine associations; France/ Austria/ Italy: Neuroscience clinics; Germany/ Luxembourg: population based cohort; Spain: GP population Total n=426 UK n=131 Italy n= 60 Spain n=107 Germany/Austria n= 83 France n=45	Total: 44.0 (+/- 11.38)	Total: F 75.0%	NA	Total: quoted as 66-100% (data NR) not quoted at item level	EUROLITE - development
Bagley et al 2012 [54]	Data from 9 countries: US, Canada, France, Germany, Spain, UK, Australia, Italy, Taiwan.	Chronic (CM) and Episodic Migraine (EM) Detailed definitions *(p410)*	International Burden of Migraine Study (IBMS) Web-based, cross-sectional population survey Participants recruited from established database of headache / migraine patients (n	Total: 40.3 (11.4) EM 40.2 (11.4)	Total: F 83.5% EM F 83.4%	N/A	Not reported at item level	MSQ 2.1 *Includes a 'review' of the development / earlier evaluation*

	Suggests that questionnaires completed in 'official language of the country' – but results then combined	EM: <15 HDPM CM: >/= 15 HDPM	63,001): all received e-invitation to 'opt in' via web-link. 30.7% responded to email invitation; 55% eligible to participate. Surveys completed by 81.9% Total n 8726 EM n 8227 CM n 499	CM 41.7 (12.1)	CM F 85.6%			*papers for MSQ and HIT-6.*
Bigal et al 2003 [69]	US US English	Patients registered with a specialist headache clinic CM: daily or near-daily headaches lasting >4hrs if untreated, >15 days per month, fulfilling CDH. EM: HIS for migraine +/- aura	Retrospective assessment of patient clinic notes (for those who had previously completed the MIDAS) CM 182 EM 86	CM 38.3 (95% CI 36.5 to 40.1) EM 36.1 (95% CI 34.1 to 38.0)	CM F 72.5% EM F 68.6%	N/A	Not reported at group or item level	MIDAS
Bjorner et al 2003a [3]	US US English	General population At least 1 headache in 4/52 prior to interview (not hangover, cold, flu)	National Survey of Headache Impact (NSHI) – longitudinal survey (baseline and 3/12) Sampling frame: randomly generated list of telephone numbers from 48 US states.	NR *Reported elsewhere in NSHI papers*	NR	N/A	N/A	HIT development paper Focus: item pool development. Informed by items from the: MSQ (v2), HDI, HIMQ, MIDAS

			Interviews with convenience sample of eligible respondents (mean duration 21.5mins (rge 17-27 mins) – schedule not detailed Telephone interviews n= 1016 Headache prevalence in 4/52 period= 45.7%					
Blumenfeld et al 2010 [71]	Data from 9 countries: US, Canada, France, Germany, Spain, UK, Australia, Italy, Taiwan. *Suggests that questionnaires completed in 'official language of the country' – but results then combined*	Chronic (CM) and Episodic Migraine (EM) Detailed definitions (p410) EM: <15 HDPM CM: >/= 15 HDPM	International Burden of Migraine Study (IBMS) Web-based, cross-sectional population survey Participants recruited from established database of headache / migraine patients (n 63,001): all received e-invitation to 'opt in' via web-link. 30.7% responded to email invitation; 55% eligible to participate. Surveys completed by 81.9% Total n 8726 EM n 8227 CM n 499	Total: 40.3 (11.4) EM 40.2 (11.4) CM 41.7 (12.1)	Total: F 83.5% EM F 83.4% CM F 85.6%	N/A	Not reported at item level	MIDAS * Sub-division of Grade IV into IV-A severe disability (score 21-40) and IV-B very severe disability (41-270): reflects number of people with CM who fall into the grade IV category. *(ceiling effect – worst scores)* *Includes a 'review' of the development / earlier evaluation papers for MSQ and HIT-6.*
Brown et al 2005 [79]	US	Care-seeking patients for migraine headache - registered at three sites representing varied	Cross-sectional survey questionnaire – primary focus evaluation of the HUI-3 in this population	44.0 (11.6)	F 87%	N/A	Not reported.	HUI-3 (MIDAS)

	US English	models of healthcare: primary care speciality clinic, non-profit HMO. EM: physician diagnosed at least 1-year before study enrolment (medical chart review)	Consecutive patients recruited (each site total n50). N 150		*(Causasian 87%)*			Mean HUI-3 scpre 0.62 (SD0.26)
Chatterton et al 2002 [40]	US US English	Migraine diagnosis At least 1 migraine per month and	Cross-sectional survey questionnaire - primary focus evaluation of the M-TAQ; two-week test-retest (of individual items) on sub-set (n 100) N 243	40.0 Range 18-63	F n 219 (91%) *(Caucasian 63%)*	N/A	Not reported.	M-TAQ – development paper (SF-36, MIDAS, Beck Depression Inventory) Test-retest at 2-weeks (n 100)
Coeytaux et al 2006 [58]	US US English	Specialist headache clinic Chronic Daily Headache (CDH): presence of headache on >/= 15 days in the month prior to enrolling in clinical trial Total recruited with CDH n 71	Randomized clinical trial: Usual medical care (UMC) (n37) v UMC plus acupuncture (n34) Questionnaires administered at baseline and 6/52. Include patient self-report of meaningful improvement/no change/ deterioration at 6/52.	Mean 46.0; range 19-83 yrs.	F 80% (n 57) (93% white (n 66))	Usual medical care (UMC) (n37) v UMC plus acupuncture (n34) (consisting of 10 treatments over 6-wks)	Complete follow-up data for 71/74 enrolled patients (96%)	HIT-6 Patient-reported change in status at 6/52: 42% improved; 44% no change; 14% worse

		Mean duration of CDH 24.2 (SD 5.8) days						
Cole et al 2009 [60]	US and Canada US English	1.Participants in clinical trial; and 2. Members of the general population EM or Self-report Headache: EM: minimum 6-mth history of Migraine (HIS criteria) with 3-12 Migraines per mth but not >15 HDPM during the 28-day prospective baseline period. Self-report headache at least once in past 4-weeks (prior to phone interview)	Retrospective data analysis of two data-sets: 1. Pooled data from 2 randomized clinical trials (RCTs) (n 916) (Topiramate for Migraine prevention) 2. Population-based database (n 1016)	1. rge 12-65 yrs 2. rge 18-65 yrs Additional data NR	NR	1. Double-blind placebo-controlled RCT 2. No intervention	Missing data: handling detailed p1181 (Bayesian multivariate imputation method)	MSQ v2.1 Focus on MID calculation for MSQ v2.1 (anchor and distribution based analysis) (p1182)
Cole et al 2007 [51]	US and Canada US English	Participants in a clinical trial	Retrospective data analysis: pooled data from 2 randomized clinical trials (RCTs) (n 916) (Topiramate for Migraine prevention)	40.7 (10.7)	NR	Double-blind placebo-controlled RCT - Prophylactic migraine treatment	Missing data detailed. MSQ v2.1: Baseline:	MSQ v2.1

		6-mth history of Migraine (IHS criteria) with 3-12 Migraines per mth but not >15 HDPM during the 28-day prospective baseline period.	MSQ v2.1 completed baseline, 2/4/6mths				range 0% (several items) to 0.44% (items 3 and 4)	Reports results for both 14-item (v2.1) and a revised 13-item measure (informed by results of initial testing) IRT evaluation
Coon et al 2012 [38]	US US English	Patients registered with specialist headache clinics (8 sites) IHS diagnosis of migraine +/- aura for at least 1 year based on medical, medication and migraine history. During 3-mths prior to study enrolment – required to have x 3 to 8 migraine attacks per month and to have used triptans at least x2 per month.	Clinical study *(before / after)* (sumatriptan and naproxen versus usual therapy) n 916 Baseline (BL) static CORS completed to reflect experience of 'usual therapy'. After 2-mths treatment with Suma/Nap they completed the EOS static CORS (re-phrased to focus on Suma/Nap); comparative CORS also completed.	44.3 (11.0); range 19-65yrs	F 87.1% *(91.2% white)*	Visit 1 (for 2-mths): participants treated any migraine with single-tablet formulation of Suma/Nap.	Not reported (at survey level)	Focus: treatment-specific measure - CORS Mean 22.4yrs (13.2; range 1-53) since Migraine onset) Average HIT-6 at baseline 61.7 (range 42-76) suggesting severe impact of headache
Cramer et al 2001 [34]	US US English	Patients with history of migraine. Migraine not defined.	Development and initial evaluation of the HANA: data from three studies:	1.Mean NR. Range 19->65	1. F 804/994 (81%) 2. F 17/28 (61%)	1. N/A 2. Migraine prophylaxis clinical trial 2. Usual care	'No floor or ceiling effects' reported (but data not illustrated)	HANA - development paper (HDI)

			1.Participants in the Life Impact Survey – a web-based survey (supported and widely publicised by various Headache and Migraine groups) N 994 2.Participants in a migraine prophylaxis clinical trial. N 28 3. 1-month test-retest cohort: 'no change in status' (anchor not reported) N 25	2. 40.7 (11.1); range 16-69 3. 44.0; range 16-62	3. F 21/25 (84%)			
Davis et al 2002 [76]	Data from 6 countries: Spain, NZ, NL, Hungary, Finland, Canada	Clinical trial participants Diagnosed according to IHS criteria N 793	Multinational, open-label trial (oral naratriptan) vs 'usual therapy for M'. Duration 3-months Baseline and 3-mths completion (end of trial)	Total population 38.4 Country mean age range: Spain 36.2 (9.1) to Canada 40.2 (8.7)	F 85% *(98% white)*			PPMQ (original version) Limited detail re PPMQ development (
Dodick et al 2007 [61]	US US English	Patients aged >18 years with chronic migraine Migraine defined according ICHD-II, with duration of 30 minutes or longer.	RCT – patients randomized 1 : 1 ratio to topiramate 100 mg/day or placebo (double-blind period 16weeks) n328	38.2	F 85.3%	N/A Report responsiveness (correlation of change scores) Calculate MID (within-person (MIC)) for MSQv2.1	NR	MIDAS MSQ v2.1 Subjective Global impression of change (SGIC)

		CM identified based on Silberstein–Lipton criteria, which required the presence of at least 15 headache days during the 28-day prospective baseline period				SGIC completed at the end of the study – 7-point scale (1= very much improved to 7= very much worse) Physician GIC also completed.		
Dowson et al 2004 [39]	International study: 5 counties: UK, US, Spain, Germany, Italy. All questionnaires translated into local language.	Patients registered at secondary care headache centres and attending for migraine treatment. Migraine diagnosed according to IHS criteria (1998; 2004): participants aged 18-65yrs; minimum 1 yr history of migraine; average 1-4 attacks per month and minimum 24-hrs between attacks, and able to distinguish migraine from other headache.	Development and initial evaluation of the M-ACT: Open, prospective, multi-national, observational, two-visit study. Baseline (n 185) 1-week test-retest (n 143) (no change in treatment during this time; but no health transition question reported). Questionnaire completed in clinic or by telephone.	44.0; range 14-87 (93% aged 18-65)	F 68% *(Caucasian 99.4%)*	N/A	NR	M-ACT – development paper (SF-36, MIDAS, M-TAQ)
Gillard et al 2012 [73]	Data from 9 countries: US, Canada, France, Germany, Spain, UK, Australia, Italy, Taiwan.	Episodic and chronic M Migraine defined according to International Classification of Headache Disorders (2nd Edition): plus	International Burden of Migraine Study (IBMS) Cross-sectional, web-based observation study: paired observations of participants Individuals randomly assigned to training or 'validation' samples	*NR in this paper for total pop.* Presented by MSQ and HIT-6 completion (Table 1 p486)	Range 83-86% female	N/A	NR – assumed how dealt with??	EQ-5D with HIT-6 and MSQv2.1 **Conclusion:** relationship between the EQ-5D and both measures is adequate to use

	Suggests that questionnaires completed in 'official language of the country' – but results then combined	Chronic >/= 15 days per month; Episodic (< 15 HD per mth) Total > 8500 CM >450 EM >8000	Aim: to develop empirical algorithms to estimate health state utility values from disease-specific QOL scores in individuals with migraine	Range (Median) 39-42 yrs (range 18-85)				regressions equations to estimate EQ-5D utility values. The preferred HIT-6 and MSQ algorithms can be used to estimate HSU in trials where a preference based measure is not used.
Kawata et al 2005 [55]	US US English	New adult patients at university headache-speciality practice Diagnosis not specified	Cross-sectional survey All new patients who presented at clinic from Jan-Sept 2001 N 309	41.0 (SD 13) Range 18-91 yrs.	F 77%	NR	Questionnaire response rate 309/369 (84%). HIT completion not reported at item level	HIT
Kilinster et al 2006 [74]	International study: 5 counties: UK, US, Spain, Germany, Italy. All questionnaires translated into local language.	Patients registered at secondary care headache centres and attending for migraine treatment. Migraine diagnosed according to IHS criteria (1998; 2004): participants aged 18-65yrs; minimum 1 yr history of migraine; average 1-4 attacks per month and minimum 24-hrs between attacks, and able to distinguish migraine from other headache.	Secondary analysis of data from the M-ACT study database (Dowson et al, 2000): evaluation of M-ACT reliability and validity. Open, prospective, multi-national, observational, two-visit study. Data analysed for total population and per country Baseline (n 185) 1-week test-retest (n 143) Questionnaire completed in clinic or by telephone.	44.0; range 14-87 (93% aged 18-65)	F 68% *(Caucasian 99.4%)*	N/A	NR	M-ACT (SF-36, MIDAS, M-TAQ)

Kimel et al 2008 [75]	US US English	Minimum 6/12 history that met HIS criteria for Migraine with aura; experienced 2-6 Migraines per mth in the 3mths prior to screening (screening assessment to confirm diagnosis)	Data from two identical phase 3 trials: fixed dose sumatriptan + naproxen sodium Vs placebo N 1304	40.1 (11.09)	F 87.8%	Fixed dose sumatriptan + naproxen sodium Vs placebo	Detailed p514-515	PPMQ-R Evaluate psychometric properties in clinical trial setting
Kosinski et al 2003 [31]	US US English	Members of the general population with experience of recent headache (self-diagnosed?) A headache in the last 4-weeks that was not due to cold, flu, head injury, hangover	Two on-line community-based surveys (platform AOL's Opinion Place). Time 1: n 1103 Time 2: n 540	Time 1. 37.0 Time 2. 37.5	Time 1: F 73% Time 2: F 72%	N/A		(HIT and) HIT-6 development paper
Lipton et al 2003 [77]	US and UK US / UK English	Migraine (n 399) and non-migraine (n 379) controls (data pooled for both populations p631) HIS defined migraine (computer-assisted telephone interview (CATI)): HIS migraine +/- aura: 6 or more M in last year, but 15 or < headaches (any type) in previous month.	Two population-based studies (UK and US) – cross-sectional survey. Evaluation of migraine, quality of life (SF-12) and depression: patients with migraine versus non-migraine counterparts.	NR in this article – web-link provided.	NR	Nil.	NR	SF-12

Lipton et al 2009 [41]	6 countries: Canada, France, Germany, Italy, Spain, USA	Community-based population IHS for Migraine +/- aura; at least x1 M per mth in the past 3mths; and no change in Rx for M in past 3mths.	Focus to establish the reliability and validity of a new measure – the M-TOQ: evaluation in five languages. 25 Primary care centres: N 253 (n50 per language) Questionnaire completion during clinic-based interview: package of measures including M-TOQ, MIDAS, HIT-6, MSQoL. *(Unclear if all pen and paper or some touch-screen completion).* Test-retest completion at 7-10 days (pen and paper completion).	43.1 (12.4)	F 90.1%	N/A Evaluation of: data quality (missing values not reported), structural validity, internal consistency reliability, item-total correlation, convergent validity.	NR	M-TOQ (HIT-6, MSQoL)
Lipton et al 2016 [59]	US US English	Eligible adults (aged 18–65 years) - International Classification of Headache Disorders (ICHD-2) diagnostic criteria for CM . To be eligible for inclusion, patients must have had >15 headache days during the 28-day screening period (baseline), during which >4 hours of each headache day were continuous headache and >50% were migraine or	Patients with CM from PREEMPT (Phase 3 REsearch Evaluating Migraine Prophylaxis Therapy) were randomized (1:1) to receive onabotulinumtoxinA or placebo for two 12-week cycles in the double-blind (DB) phase, followed by three 12-week cycles of open-label (OL) onabotulinumtoxinA (onabotulinumtoxinA/onabotulinumtoxinA (O/O) and placebo/onabotulinumtoxinA (P/O) groups, respectively). n1236 participants (O/O, n607; P/O, n629) participated in both phases HRQoL endpoints were assessed over 56 weeks using the HIT-6 and MSQv2.1	Baseline 41 (SD 10) yrs White 90%	F 85%	PREEMPT clinical trial Focus of report – the pooled HRQOL outcomes for 56-week treatment period. Baseline Mean HIT-6 65.4 (4.2) Mean MSQ: RR 38.6 RP 56.0 EF 42.2	Missing HIT-6 data were imputed using a prespecified, modified last observation carried forward technique (6,7). All observed MSQ data were analyzed without imputation for missing values.	HIT-6 36-49 no impact 55-55 substantial impact 60-78 severe impact MSQv2.1 Range 0 (poor) to 100 (goodHRQoL) *Paper describes where the measures have exceeded*

		probable migraine days. Mean duration since CM onset 19yrs				At baseline most patients were severely debilitated by their migraines: 93% reporting a total HIT-6 score >60 (severe impact) and another 5% reporting a score of 56–59 (substantial impact)		*proposed MIC or MID*
Magnusson et al 2012 [32]	Canada US English	Pre-IHS definition for chronic migraine; patients fulfilled criteria for transformed migraine +/- medication overuse All patients were diagnosed with Migraine and had headache on >15 days per mth (fulfilling IHS criteria)	Patient registered with specialist headache centre studies: 1) Canadian Headache Outpatient Registry and Database (CHORD) (n 83); 2) Calgary Specialist headache clinic study (n 76) Groups had similar demographics, number of H days and amount of H-related disability	1)41.0 2) 41.2	1) F n64; M n19 2) F n64; M n12	N/A	N/R	SF-36 Headache Modification (SF-36, HIT-6) Group 1 completed the SF-36 Headache Modification; Group 2 the original SF-36 (v1)
Martin et al 2000 [36]	US US English	Patients attending 4 outpatient headache speciality clinics Diagnosed according to IHS criteria	Multicenter, nondrug, prospective, parallel group, quasi-experimental design. N= 267 (157 new and 110 stable)	New patients: 39.0 Stable patients: 44.6	F 90.6%	Stable patients: TAU New patients: acute or prophylactic medications recommended by headache specialist	Item level: no missing observations or out of range values for either group	MSQ v2.1

Martin et al 2008 [65]	US US English	Sub-set of participants with CH in a large multi-site RCT. RCT inclusion: >21 yrs old; MIDAS score >5; intending to continue with general medical care; CH diagnosis defined below: CH: tension-type, migraine or mixed aetiology. Diagnosed by primary care physician (frequent and/or difficult to manage headaches)	Secondary analysis of RCT data: Headache Management Progammes (HMP): aimed at reducing H-related disability, improving process of care, reducing management costs. Total n 124 Primary focus: MTSM evaluation in a sub-set of RCT population; Baseline (parts 1 and 2) and 6-mths (parts 3 and 4): self-completion (mail)	45.4 (11.6)	F 75% *(59.7% Caucasian)*	N/A	NR	MTSM (SF-36, MIDAS)
McKenna et al 1998 [37]	UK and US UK and US English Part of international study conducted in 8 countries.	Patients with history of migraine. UK: Migraine diagnosed by clinician US: Migraine diagnosed by clinical specialist No further definition provided.	Development and initial evaluation of UK version. 1.Qualitative: interviews (UK n 30; US n25) and focus groups (US). Recruitment: UK from general practice, British Migraine Association, pharma company employees. US: from specialist clinics. 2. Postal survey (UK): Baseline n 87/90 completed questionnaire; 2-week test-retest questionnaire n87/87.	1. NR 2. 47.6; range 22-92	1.NR 2. F n72 (83%)	NA	NR	MSQoL – development paper

Niere & Quin 2009 [30]	Australia Australian English	Patients attending private practice out-patient physiotherapy clinics (n45) for headache management IHS diagnostic criteria (1988) for Migraine +/- aura, Migraine with aura and tension type headache. Diagnostic criteria for cervicogenic headache.	Cross-sectional evaluation of the HDQ Clinicians' selected consecutive patients meeting inclusion criteria N 111	38.3 (12.2) Range 18.0-74.0	F n93 (83.8%)	NA	Item response rates: 95%-100% (14 of 16 items at least 98%)	HDQ – development paper
Patrick et al 2000 [52]	8 countries: US, UK, France, Denmark, Germany, Italy, Spain, Sweden.	Previous participants in placebo-controlled trial: recruited into long-term observational study. Registered with specialist headache clinics. IHS – diagnosed with Migraine	Non-comparative long-term observational study of zolmitriptan (Zomig) for acute treatment of migraine attack of any intensity (over 12-mths) N 1383	41.2 (10.01); range 12-66	1190 F; 193 M (86:14) Caucasian: 96.5%			Focus: MSQoL (20-item) (Also SF-36 (US only; n= 1115). Questionnaires completed: after treating 1, 5 and 17 M attacks or at 3-mthly intervals
Patrick et al 2003 [42]	US and UK English	Development and initial evaluation of new measure (MTSM) Participants identified by their referring clinician as "a migraine patient starting a new treatment."	1.Participants in initial interviews / focus groups – item generation (US (30); UK (24)) and confirmation (23); Headache experts (US 3; UK 1). No additional detail. 2.Participants in clinic-based study (n=29) – to test the feasibility of using the MTSM in a clinical setting and to generate a preliminary data set from small group of patients.			New migraine treatment. 22 (75.9%) started on a triptan (ie, sumatriptan, naratriptan, rizatriptan, zolmitriptan, or almotriptan); 12/22 (54.5%) were already receiving a triptan, and the other 10	NR	Development and initial evaluation of the MTSM Questionnaire completed baseline and follow-up.

				42.0 (11.3); range 24-64	28 F (97%) Caucasian 25 (86%)	(45.5%) were receiving an analgesic. Remaining 7 (24.1%) were changing migraine drugs; of these, 2 participants had been receiving a triptan: 1 was now starting an analgesic (naproxen) and 1 was starting an antidepressant (citalopram). The 4 participants who were originally taking an analgesic started a new analgesic (naproxen or butalbital). The remaining participant had been receiving an anxiolytic (alprazolam) and was now starting paroxetine (an antidepressant/anxiolytic also used for headaches).		Initial evidence for validity assessed against 24-hr MQoLQ and SF-36.
Pathak et al 2005 [33]	US and Germany	Migraine (IHS): at least 3 migraines in previous 12mths.	Stages in the development of the FAIM – a new measure. 1) Focus groups in US and German – limited detail 2) Pre-testing – item evaluation and reduction: n153 US and n148 Germany 3) Pilot test and final item reduction: n75 US and n83 Germany	Study 2: US 37.5 (13.2) G 43.3 (13.7) Study 3: US 38.8 (11.2) G 41.0 (13.5)	F Study 2: US 66.2% G 71.5% Study 3: US 76.8 % G 69.9%	Stages in PROM development. No treatment.	Detailed per stage of developmet	Development of the FAIM
Rendas-Baum et al 2013 [12]	US, Canada, UK, Croatia, Germany and Switzerland.	International trial participants with diagnosis of migraine:	Secondary analysis of data from two multicentre double blind placebo controlled RCTs of chronic migraine patients receiving BOTOX as prophylaxis	Study 1: 41.6 (10.5)	Trial 1: F 87.5%		Not reported	Comparative evaluation of the MSQ v2.1 and HIT-6

		ICHD-II for migraine, with the exception of "complicated migraine" ≥15 headache days during 4-week baseline phase; each headache day of ≥4 hours of continuous headache; ≥50% of baseline headache days migraine/probable migraine days.	Total N 1376 Trial 1: n 672 Trial 2: n 704	Study 2: 41.0 (10.6)	Trial 2: F 85.4%			
Rendas-Baum et al 2014 [53]	US, Canada, UK, Croatia, Germany and Switzerland.	International trial participants with diagnosis of migraine: ICHD-II for migraine, with the exception of "complicated migraine" ≥15 headache days during 4-week baseline phase; each headache day of ≥4 hours of continuous headache; ≥50% of baseline headache days migraine/probable migraine days.	Secondary analysis of data from two multicentre double blind placebo controlled RCTs of chronic migraine patients receiving BOTOX as prophylaxis Total N 1376 Trial 1: n 672 Trial 2: n 704	Study 1: 41.6 (10.5) Study 2: 41.0 (10.6)	Trial 1: F 87.5% Trial 2: F 85.4%		Not reported	Comparative evaluation of the HIT-6 and MSQ v2.1
Revicki et al 2006 [43]	US US English	Primary Care and neurology speciality clinics (n=50)	Longitudinal observational study: patients receive usual medical care; study investigators had discretion to change or prescribe medications	39.0 (11.0)	F n 181 (91%)		Detailed Table 3 (p 246)	PPMQ-R Development / revision paper

		Documented diagnosis of migraine +/- aura (1988 IHS criteria 1.1 and 1.2): 2 to 8 Migraine attacks per month for at least 3-mths prior to study enrolment. Able to distinguish Migraine from other Headache	N 200 Convenience sampling from participating clinics (not detailed)		*(Caucasian n 63 (82%))*		Ceiling effects for all 10 Bothersome items (>50%) and two ease of use items.	Participants kept a Migriane diary and completed the draft PPMQ-R at 24-hrs post-Rx for each M attack (MSQv2.1)
Sauro et al 2010 [70]	Canada Canadian English	Patients registered with Neurology Outpatient practices - patients referred by family physician or other specialist Patients diagnosed according to IHS criteria - except patients with chronic daily headache (headache on 15 days a month or more).	Cross-sectional evaluation Patients identified from the Canadian Headache Outpatient Registry and Database (CHORD) – patients registered with five neurology outpatient practices (Sept 2001-Jan 2004) N 798	40.3 (SD 13.7)	F 77%	NR	Response rate 92% Item level: only those with 'valid scores' included in the analysis.	HIT-6
Sieber et al 2000 [64]	Canada Canadian English	n89 adults 'known to suffer with migraine'	Cross-sectional, comparative evaluation of QWB and QWB-SA for patients with migraine. Postal self-completion of QWB-SA Telephone-administered completion of QWB	42.2 (9.8) Range 36 to 64yrs	F 87%	Questionnaires completed at 3-points: first, on a day when migraine had not been experienced within previous 7-days; 2nd and 3rd within 48hrs of onset of migraine.	Not reported. Greater number of completions of QWB than QWB-SA	QWB (interview) QWB-SA (self-administered)

Stafford et al 2012 [72]	UK English	Members of the general population who had recently experienced a Migraine: IHS definition: Migraine +/- aura. At least 1 Migraine in the last 7 days and history of physician diagnosed Migraine for at least 6mths.	Cross-sectional observational study. Recruited via Migraine support groups and support group databases – eligibility confirmed via telephone interview. Aged >18yrs. N 105 Mean number of monthly Migraines 5.22 (4.1); range 1 to 20. MIDAS grade: iv 51/106 Iii 32/106 Ii 14/106 I 9/106	47.45 (11.71)	F n 81 (76.4%) *(Caucasian n 89 (83.2%))*		Missing items dealt with as recommended by EQ-5D developers 1 patient did not complete EQ-5D for current health status – therefore numbers reduced from 106 to 105	EQ-5D (3D) Focus: generating utilities to reflect Migraine severity for participants most recent attack (past 4-weeks) and their current health state outside of an attack. Tension-related aspects of M not evaluated. EQ-5D self-completed retrospectively to reflect most recent M: completed for each level of M severity experienced (mild, mod and/or severe) during this attack. Also completed EQ-5D to reflect current non-M health state.
Stewart et al 1999 [35]	US and UK US and UK English	Computer-Assisted Telephone Interview (CATI) based diagnosis of migraine (IHS criteria)	Postal self-completion of MIDAS – population based samples of patients with migraine-headache confirmed with CATI: UK n100 and US n97 Baseline and test-retest completion at 18-days	Range 18-55yrs	F US 83.5% UK 60%	N/A Evaluation of MIDAS reliability.	NR	Initial development / refinement of MIDAS and testing of reliability

Stewart et al 1999 [66]	US US English	Computer-Assisted Telephone Interview (CATI) based diagnosis of migraine (IHS criteria)	Postal self-completion of MIDAS – population based samples of patients with migraine-headache confirmed with CATI: US n97 and n80 non-migraine subjects Baseline and test-retest completion at 21-days	Range 18-55yrs	F US 83.5%	N/A Evaluation of MIDAS reliability	NR	Initial reliability resting of MIDAS
Stewart et al 2000 [67]	US US English	Patients with a confirmed diagnosis of migraine Initial migraine status determined from IHS based CATI algorithm Migraine status confirmed by clinical examination at initial clinic visit	Population based interview survey n=144 *(a total of 12967 diary days)*	37.6 (9.3)	F 75.7%	NA (Treatment As Usual)	Interview participation rate 67% Inadequate diary entry 16.5%	MIDAS
Stewart et al 2003 [68]	US and UK English	Total n= 397 59% MIDAS grade III or IV Initial migraine status determined from IHS based CATI algorithm	Secondary data analysis from three population based studies	Mean NR 62% aged 25 – 44yrs (range <25 to 55+)	F 78%	N/A	NR	Focus: relationship between headache features (freq, pain intensity, quality, assoc symptoms) and the MIDAS Telephone interview-based completion??

Turner-Bowker et al 2003 [7]	US US English	General population Patient self-report or patient-reported doctors' diagnosis of migraine	Cross-sectional Internet (AOL) and mail population survey Convenience population sample Total N=7557 Migraine n=1478	Range 35–44 years	F 53%	Not assessed	Total questionnaire response rate 27.8% Internet response rate unreported	SF-8
Vernon & Lawson 2015 [29]	Canada/UK Canadian English	Patients attending a chiropractic clinic and self-diagnosed with primary migraine, tension-type or cervicogenic headache	Cross-sectional completion of HADLI Participants recruited by advertisement/ personal solicitation N 53	37.3 (12) yrs.	F 41 (77.4%)	NA	NR	HADLI development paper
Ware et al 2003 [57]	US US English	General population At least 1 headache in 4/52 prior to interview (not hangover, cold, flu)	Two studies: 1.National Survey of Headache Impact (NSHI) – longitudinal survey (baseline and 3/12) Sampling frame: randomly generated list of telephone numbers from 48 US states. Telephone interviews with convenience sample of eligible respondents n= 1016 Headache prevalence in 4/52 period= 45.7% 2.On-line community-based survey (platform AOL's Opinion Place): respondents randomly			Detail: Ware 2000 *Med Care) and Bjorner 2003a [9]		HIT (total) CAT-HIT HIT-6-D ('static') and improved version (HIT-6) in study 2 SF-8 Focus: evaluation of CAT-based estimates vs 'static' scores: reliability and validity (plus

			selected and screened for eligibility: n 1103 included. Test-retest at 2-weeks for sub-group (n 540).					respondent burden)
Xu et al 2011 [78]**	US US English	Participants in a multi-centre clinical trial. M defined by IHS (+/- aura); Between 1 – 6 moderate to severe Migraine attacks per month	Secondary analysis of data from a multi-centre (20 sites) double-blind RCT of treatment for acute Migraine (Talcagepant). Rx duration was for a single Migraine attack of moderate to severe intensity. N 330 All data pooled for EQ-5D analysis. Disutilities calculated for selected patients with mod/severe Migraine pain at baseline who reported pain freedom at 24-hrs (difference in EQ-5D scores between time-points calculated)	Mean NR Range 20-65 yrs.	F n 292 (88.5%) *(Caucasian n 259 (78.5%))*	.	NR	EQ-5D (3D) Completed at baseline (whilst experiencing a mod/severe M and prior to dosing) and 24-hr post Rx within an acute M attack. Patients also completed Pain levels during this time (4-grade: no/mild/mod/ severe).
Yang et al 2010 [56]	US US English	Population based survey – self-completed questionnaires CM and EM: Categorised into 3-groups: 1) CM (=/>15 HDPM) = 6.4%;2) EM (<15 HDPM) = 42.1%; 3)	Participants in National Survey of Headache impact (NSHI) (n= 1096) and the HIT-6 Validation study (n= 54) N= 2049 Focus of analysis: reliability and validity of HIT-6 in patients with varying headache frequency days.	Mean NR Range 18-69yrs 56.3% aged 18-39	F 75%	NA	NR	HIT-6 (SF-8) HIT-6 scores calculated across the 3 groups: 1) CM = 62.5 (7.8); 2) EM = 60.2 (6.8); 3) Non-M headache = 49.1 (8.7)

		Non-M headache = 51.5% Survey included a Migraine screener ID Migraine criteria – p359) and number of headache days per month (HDPM)						
Study **(Author;yr)** **[ref]**	**Country[g]** **Language[g]**	**Population and Headache definition**	**Study: design, setting, sample size[g]**	**Mean [g]age (SD; range)**	**Gender distribution (F Female)[g]**	**Treatment descriptions[g]**	**% missing responses: acceptable?[g] ***	**PROM focus;** **Additional info [g]**

Footnote:

- Missing values: a. At survey level; b. At item level (data quality – frequency with which items were missing (ie. Non-completed) and how did the authors deal with missing values?

Bagley 2012: HDPM Headache days per month (p410 for detail); IBMS – International Burden of Migraine Study (2009).

Appendix 6: Table 5 Content comparison at item level (number of items) of condition-specific measures (n= 17)

PROM[a]	Items (n) [b]	Domains of Health-related Quality of Life (Ferrans et al, 2005 [25])*												Response to Treatment			
		Symptom Status				*Functional Status*							*General Health Perception*				
		Symptoms				Physical	Social / Role				Psychological						
							Social / Role		ADL / Work		EWB	Cognition		Return to normal	Confidence / Satisfaction with Rx	Side effects	Other
		General	Headache Frequency	Headache Intensity	Pain	Physical	Limit	Prevent	Limit	Prevent							
Condition-specific (17)																	
Migraine impact (5/17)																	
FAIM [33]	9 + 5						5				5		4				
HANA [34]	7	1					1		1		3		1				
MIDAS [35]	5 + 2		1*	1*				1	4								
MSQv2.1[36]	14	2					2	1	3	2	3	1					
MSQoL [37]	20						6		10		4						
Headache impact (6/17)																	
EUROLIGHT [d] [28]	103	X	X	X	X	X	X		X		X	X					
HADLI [29]	9	1				1		2	4			1					
HDQ [30]	9				2		1		3	3							
CAT-HIT [e] [3,57]	54	X			X		X		X		X	X					
HIT-6 [31]	6	1			1		1		1		1	1					
SF-36 Headache [32]	6/36				2	1	2		1								
Response to migraine-specific treatment (6/17)																	
Static CORS Comp CORS [38]	24 8	9 4			5				2		3	1		1 1	3 3		
M-ACT [39]	27 4													20 2	7 2		
M-TAQ [40]	9	1	1							1				1	1		4
M-TOQ15 M-TOQ-5 [41]	15 5	1		1	2 1				2 1		1			1	4 2	3 1	
MTSM [42]	45	1			3				1						3		
PPMQ-R [43]	19 10	3 8			6		2		2		1 1	 1	2 (cost)		3		

Footnote:

****Ferrans et al (2005) revision to the Wilson and Cleary Model of Health Related Quality of Life (HRQOL)[25]:*** The model describes five levels of patient outcomes from biological and psychological variables through to overall quality of life (subjective well-being assessed by an individuals perceived level of happiness or satisfaction with life), and includes symptoms (for example, physical, emotional, cognitive symptoms perceived by the patient), functional status (for example, physical, social, role and psychological function) and general health perceptions (a subjective rating incorporating all of the preceding health concepts). Additionally, characteristics of the individual (such as values and preferences) and those of the environment (such as social, economic and psychological support) are considered.

[a] PROM content; [b] Number of items per PROM

Migraine-impact:

[a] FAIM – Functional Assessment in Migraine: 9 items across three domains. 1) Attention/Thought (5 items); 2) Perception (4 items); 3) Activity and Participation (5 items)

HANA - Headache Needs Assessment: 7 items to reflect migraine frequency and bothersomeness. 1 item in each of following area: Anxiety/worry; depression/discouragement; self-control; energy; function/work; family/social activities; overall impact of migraines.

MIDAS - Migraine Disability Assessment Score: 3 domains (5 scored items): Missed days/ reduced productivity at paid work (2 items); Missed days/ reduced productivity at household work (2 items); Missed non-work activities (1 items)

MSQv2.1 - Migraine-Specific Quality of Life Questionnaire version 2.1: 3 domains (14 items): 1) Role function – Restrictive (RR)(7 items): social – family/friends; leisure (2); work/ADL (3); cognition (1); symptoms: fatigue (1); 2) Role function – Preventive (RP)(4 items): symptoms: fatigue (1); work /ADL (2); social (1); 3) Emotional function (EF)(3 items): frustration (1); feeling a burden (1); letting others down (1).

MSQOL - Migraine-Specific Quality of life: 3 domains (20 items): 1) Avoidance behaviours (10); 2) Social relations (6); 3) Feelings (4).

Headache impact:

EUROQLIGHT: 6 sections (103 items): 1) Biographical (age, gender, language and employment); 2) Screening questions for headache; 3) Diagnostic questions; 4) Questions about any headache experienced 'yesterday' (point prevalence); 5) Use of healthcare resources (medicines, investigations, consultations, etc.); 6) Impact of headache on work, family life and social activities: includes items taken from the WHOQOL (8-items), the HALT-index and HADS. [d] *EUROLIGHT – number of items per domain not clear (denoted by X).*

HADLI - Headache Activities of Daily Living Index: 1 domain (9 items): Personal care, Lifting, Reading (including computers), Sleeping (over last week), Exercising (over last week), Social activities, Work, Driving or travelling, Recreation.

HDQ – Headache Disability Questionnaire: 3 domains (9 items): 1) Pain (2 items): usual pain intensity; when pain is severe; 2) Activity Limitation (4 items): Decreased efficiency in non-work activities; Decreased ability to work/study; Decreased efficiency in housework or chores; Proportion of times when work is missed; 3) Activity Prevention (3 items): Number of days where chores prevented;
Number of days non-work activities prevented; Number of days in last month when had to lie down for >1 hour.

HIT (CAT-HIT) - Headache Impact Test: One domain 'Headache Impact': 54-item 'item bank'. Items cover a wide spectrum of headache impact, including minor headache. Items cover of pain, role and social functioning, energy/fatigue, cognitive function, and mental health.

[e] CAT-HIT: Number of completed items is individualised and determined by 'stopping rule': mean number 6/54 items (denoted by X)

HIT-6 – Headache Impact Test -item (static): 6 domains (6 items): 1) Headache pain severity(1); 2) Social functioning - u*sual daily activities*(1); 3) Role functioning *(how often do you wish you could lie down?)* (1); 4) Vitality(1); 5) Cognitive functioning(1); 6) Psychological distress(1).

SF-36 Headache - Inserts 'including your headaches' to 6/36 items: physical functioning (item 3); role limitation - physical (item 4); social functioning (items 6 and 10); bodily pain (items 7 and 8)

Response to migraine-specific treatment:

CORS - Completeness of Response to Migraine Therapy Survey: Two modules:1.'Static CORS' – 5 domains (24 items): 24-items address: Symptoms, Emotional well-being, Cognition, ADL, Function, Confidence in current medical. 2. 'Comparative CORS (8 domains (8 items)): Completeness of Relief (1), Speed of Relief (1), Persistence of Relief (1), Return to Normal Function (1), Fatigue (1)
Confidence in Treatment (2), Overall Satisfaction (1) – most satisfied.

M-ACT - Migraine – Assessment of Current Therapy: 4 domains (27 long-form/ 4 short): Headache impact (11/1); Global assessment of relief (9/1); Consistency of response (3/1); Emotional response (4/1).

M-TAQ - Migraine Therapy Assessment Questionnaire: 9 items - Migraine control, frequency of attacks, knowledge and behavioural barriers, economic burden, treatment satisfaction.

M-TOQ - Migraine-Treatment Optimization Questionnaire M-TOQ: 15- and 5-item versions: 5 domains (15/5 items): Functioning (3/1), Rapid relief of headache (3/1), Consistency of response (3/1), Prevention of recurrence (3/1), Side effects (3/1).

MTSM - Migraine Treatment Satisfaction Measure: 4-part assessment: 1) Expectations of Treatment for Migraine (TE-M; 9 items); 2) Importance Ranking for Migraine Treatment (IR-M; 9 items); 3) Outcomes of Treatment for Migraine (TO-M; 9-items); 4) Satisfaction with Migraine Treatment (PST-M; 9-items). Considered across 9 attributes ('items') associated with migraine relief: Pain relief; Speed of relief; Freedom from pain; Additional symptoms; Confidence in treatment; Disruption in life; Dosing; Freedom from relapse; Ease of use

PPMQ-R - Patient Perception of Migraine Questionnaire -Revised: Core: 4 domains (19 items): Efficacy: satisfaction with treatment efficacy (11); Function: ability to perform usual activities (4); Ease of Use (2); Cost (2) (*may be removed where 'cost' is not a consideration*). Additional fifth domain (10 items): 'Bothersomeness' or 'Tolerability' - related to side effects (10). 3 global items: Overall satisfaction with medication effectiveness; Side effects; General treatment.

Appendix 7: Table 6: Methodological quality (COSMIN[a]) of each study (n=46) per PROM (n=23) and investigated measurement property.

PROM[b] Study (n)	Country (language)	Headache definition[c]	(n)	Reliability			Validity				Responsiveness		Interpretability
				Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	*Known-groups validity*	Responsiveness (COSMIN)	Responsiveness - other	
Condition-specific (17)													
Migraine-impact (5/17)													
FAIM (1)													
Pathak et al 2005 [33]	US English German	M	69 83	Good			Fair	Good	Good				
HANA (1)													
Cramer et al 2000 [34]	US English	M		Poor	Poor		Poor	Poor	Poor	*Poor*			
MIDAS (12)													
Stewart et al 1999a [35]	US and UK English	M (IHS)	US 97 UK 100	Poor	Poor								
Stewart 1999b [66]	US English	M (IHS)	97	Poor	Poor								
Stewart et al 2000 [67]	US English	M (IHS)	144						Fair				
Stewart et al 2003 [68]	US and UK English	M (IHS)	397						Poor				
Bigal et al 2003 [69]	US English	CM	182							*Fair*			
Dodick et al 2007 [61]	US English	CM (ICHD-II)	328								Fair		
Martin et al 2008 [65]	US English	CH – TT/M/mixed	124						Fair				
Lipton et al 2009 [41]	Multiple	M (IHS)	253	Poor					Fair				
Sauro et al 2010 [70]	Canadian English	CH (IHS)	798						Fair	*Fair*			
Blumenfeld et al 2010 [71]	Multiple	CM; EM	8726							*Fair*			
Yang et al 2010 [56]	US English	CM; EM	>600						Good	*Good*			
Bagley et al 2012 [54]	Multiple	CM; EM	8726 total						Fair				
Stafford et al 2012 [72]	UK English	M (IHS)	105						Fair				
MSQ v2.1 (9)													
Martin et al 2000 [36]	US English	M (IHS) / EM	267	Good	Good		Fair		Excellent	*Excellent*			

Revicki et al 2006 [43]	US English	EM (IHS 1.1, 1.2)	200						Good				
Dodick et al 2007 [61]	US English	CM (ICHD-II)	328								Fair		Poor
Cole et al 2007 [51]	US English	EM (IHS)	916	Excellent				Excellent	Fair				
Cole et al 2009 [60]	US English	EM											Fair
Blumenfeld et al 2010 [71]	US English	CM and EM								*Fair*			
Bagley et al 2012 [54]	US English	CM; EM	8726 total	Fair					Fair	*Fair*			
Gillard et al 2012 [73]	Multiple	CM; EM	8726							*Fair*			
Rendas-Baum et al 2013 [12]	US English	CM (ICHD-II)	1376	Good	Good			Good	Good	*Good*		*SRM*	
Rendas-Baum et al 2014 [53]	US English	CM (ICHD-II)	1376						Good				
Lipton et al 2016 [59]	US English	CM (ICHD-II)	1236									*% exceeding MIC or MID*	
MSQoL (3)													
McKenna et al 1998 [37]	UK English	M (IHS)		Poor	Fair		Excellent		Poor	*Fair*			
Patrick et al 2000 [52]	US English	M	1376	Fair			Fair	Good		*Fair*		*ES; SRM*	Poor
Lipton et al 2009 [41]	Multiple	M (IHS) (EM?)	253	Poor					Fair				
Headache-impact (6/17)													
EUROLIGHT (1)													
Andree et al 2010 [28]	Multiple UK English	Headache – all types	426 131	Poor	Poor		Fair		Poor	*Poor*			
HADLI (1)													
Vernon et al 2015 [29]	Canadian English	Self-diagnosed headache: M, TT, CG	53	Fair			Poor	Fair					
HDQ (1)													
Niere & Quin 2009 [30]	Australian	M (IHSD); Mixed: TT, CG, 'other'	111	Fair			Poor	Fair					

HIT (3)													
Bjorner et al 2003a [3]	US English	H (C or E)	1016				Fair	Excellent					
Ware et al 2003 [57]	US English	H	1.1016 2.1103		Fair		Fair		Good	*Good*	Poor		
Kosinski et al 2003a [31]	US English	H	1103		Fair		Fair		Fair	*Fair*			
HIT-6 (12)													
Kosinski et al 2003a [31]	US English	H	1103	Fair	Fair		Fair		Fair	*Fair*	*Fair*		
Ware et al 2003 [57]	US English	H	1.1016 2.1103		Fair		Fair		Good	*Good*	Poor		
Kawata et al 2005 [55]	US English	H	309	Fair				Fair	Fair				
Coeytaux et al 2006 [58]	US English	CDH	71										Fair
Lipton et al 2009 [41]	Multiple	M (IHS) (EM?)	253	Poor					Fair				
Sauro et al 2010 [70]	Canadian English	CH (IHS)	798						Fair	*Fair*			
Yang et al 2010 [56]	US English	CM; EM	>600	Good	Fair				Good+	*Good*			
Gillard et al 2012 [73]	Multiple	M (ICHD): CM and EM	9048							*Fair*			
Bagley et al 2012 [54]	US English	CM; EM	8726 total						Fair				
Magnusson et al 2012 [32]	US English	CM (+/- MOU)	159						Poor				
Rendas-Baum et al 2013 [12]	US English	CM (ICHD-II)	1376						Good				
Rendas-Baum et al 2014 [53]	US English	CM (ICHD-II)	1376	Good	Good			Good	Good	*Good*		*SRM*	
Lipton et al 2016 [59]	US English	CM (ICHD-II)	1236									*% exceeding MIC or MID*	
SF-36 'Headache-specific' (1)													
Magnusson et al 2012 [32]	US English	CM	159				Poor / nil		Fair				
Response to migraine-specific treatment (6/17)													

CORS (1)													
Coon et al 2012 [38]	US English	M (IHS)	916	Fair			Good	Fair	Fair		Fair		
M-ACT (2)													
Dowson et al 2004 [39]	Multiple	M (IHS)	185		Fair		Poor		Poor				
Kilinster et al 2006 [74]	Multiple	M (IHS)	185		Fair				Poor				
M-TAQ (1)													
Chatterton et al 2002 [40]	US English	M (IHS)	243		Fair		Fair		Fair				
M-TOQ 19/15/5 (1)													
Lipton et al 2009 [41]	Multiple	M (IHS) (EM?)	253	Fair	Fair		Fair	Fair	Fair				
MTSM (2)													
Patrick et al 2003 [42]	US English	M	29	Poor			Good		Fair	*Fair*			
Martin et al 2008 [65]	US English	CH – TT/M/mixed	124	Fair			Poor	Fair	Good	*Good*			
PPMQ-R (2)													
Revicki et al 2006 [43]	US English	EM (IHS 1.1, 1.2)	200	Excellent	Fair		Excellent	Excellent	Excellent	*Good*		*ES*	Fair
Kimel et al 2008 [75]	US English	M (IHS)	1304	Excellent				Good	Good	*Good*			
Generic measures (6)													
Profile measures (3/6)													
SF-36 (5)													
Patrick et al 2000 [52]	US English	M	1376									*ES, SRM*	
Martin et al 2000 [36]	US English	M (IHS) / EM	267						Good				
Davis et al 2002 [76]	Multiple	M (IHS)	793						Poor				
Kawata et al 2005 [55]	US English	H	309						Fair				
Martin et al 2008 [65]	US English	CH – TT/M/mixed	124						Fair				
SF-12 (1)													
Lipton et al 2003 [75]	US and UK English	M (IHS)	M 399							*Fair*			

			Control 379										
SF-8 (4)													
Kosinski et al 2003a [31]	US English	M (self-report)	1103						Fair				
Ware et al 2003 [57]	US English	H	1.1016 2.1103						Good				
Turner-Bowker et al 2003 [7]	US English	M (self-report)	465						Good	*Good*			
Yang et al 2010 [56]	US English	CM; EM	>600						Good	*Good*			
Utility measures (3/6)													
EuroQoL EQ-5D-3L (3)													
Xu et al 2011 [78]**	US and UK English	M (HIS)	330							*Poor*			
Gillard et al 2012 [73]	Multiple	M (ICHD): CM and EM	9715 CM 555 EM 9160							*Fair*			
Stafford et al 2012 [72] *	Uk English	M (IHS)	105						Fair	*Fair*			
HUI-3 (1)													
Brown et al 2008 [79]	US English	EM	150							*Poor*			
QWB and QWB-SA (1)													
Sieber et al 2000 [64]	US English	M	89		Poor					*Fair*			

Footnote:

[a] COSMIN – Consensus on Standards for Measurement Instruments. Four-grade rating for study methodological quality: Excellent, Good, Fair Poor. [20,21]

[b] PROM acronyms (detailed in text and Tables 1 and 2)

[c] Headache definition: H – Headache (general); CH – Chronic Headache; TT – Tension Type; CG – Cervico Genic; CDH – Chronic Daily Headache; M – Migraine; EM – Episodic M; CM – Chronic M; IHS – International Headache Society – headache classification system (https://www.ichd-3.org/); ICHD – International Classification of Headache Disability (https://www.ichd-3.org/).

*Focus of paper: generating utilities to reflect Migraine severity; **Focus of paper: calculating disutilities.

Table 1: Data synthesis, levels of evidence and overall quality of reviewed PROMs (n=23)[a]

PROM[b] / Study	Number of evaluations	Reliability			Validity		Construct Validity		Responsiveness	Interpretation
		Internal consistency	Temporal stability	Measurement error	Content validity	Structural validity	Hypothesis testing	Known-groups	Responsiveness	
Condition-specific (17)										
Migraine- impact (5/17)										
FAIM [33]	1	+ Moderate			? Limited	+ Moderate	+ Moderate			
HANA [34]	1	+ Unknown	+ Unknown		? Unknown	? Unknown	? Unknown	? Unknown	ES only	
MIDAS [35]	13	+ Unknown	+ Unknown				+ Moderate	+ Moderate		
MSQ v2.1 [36]	11	+ Strong	+/- Conflicting		+ Limited	+ Strong	+ Moderate	+ Moderate		+ Limited
MSQoL [37]	3	+ Limited			+ Strong	+ Moderate	? Unknown	? Unknown		? Unknown
Headache-impact (6/11)										
EUROLIGHT [28]	1	? Unknown	? Unknown		? Limited		? Unknown	? Unknown		
HADLI [29]	1	+ Limited			? Unknown	+ Limited				
HDQ [30]	1	+ Limited			? Unknown	+ Limited				
HIT [3,57]	3		+ Moderate		+ Limited	+ Moderate	+ Moderate	+ Moderate		
HIT-6 [31]	13	+ Strong	+ Moderate		+ Limited	+ Moderate	+ Strong	+ Strong		+ Limited
SF-36 'Headache' Modification [32]	1				? Unknown		? Unknown			
Response to migraine-specific treatment (6/17)										
CORS [38]	1	+ Limited			+ Moderate	+ Limited	? Limited		+ Limited	

M-ACT [39]	2		+ Moderate		? Unknown		? Unknown			
M-TAQ [40]	1		+ Limited		+ Limited		+ Limited			
M-TOQ [41]	1	+ Limited	+ Limited		+ Limited	+ Limited	+ Limited			
MTSM[42]	2	+ Limited			+ Moderate	+ Limited	+ Moderate	+ Moderate		
PPMQ-R [43]	2	+ Strong	+ Limited		+ Strong	+ Strong	+ Strong	+ Strong		+ Limited
Generic measures (6)										
Profile measures (3/6)										
SF-36 [44]	5						+ Moderate			
SF-12 [45]	1							+ Limited		
SF-8 [46]	4						+ Moderate	+ Moderate		
Utility measures (3/6)										
EuroQoL EQ-5D-3L [47]	3						+ Limited	+ Limited		
HUI-3 [48]	1							? Unknown		
QWB / QWB-SA [49,50]	1		+/- Conflicting					? Unknown		

Footnote: [a] Data synthesis: The data were qualitatively synthesized to determine the overall quality of measurement properties and acceptability of each reviewed PROM. The synthesis took the following factors into account: 1) methodological quality of the reviewed studies (COSMIN scores); 2) the number of studies reporting evidence of measurement properties per PROM; 3) the results for each measurement property for each PROM; and 4) the consistency of results between reviewed studies.

The data synthesis score has two elements [19,27]:

First, the overall quality of a measurement property was reported as: adequate (+), not adequate (-), conflicting (+/-), or unclear/indeterminate (?) (Table 1 for detail).

Second, levels of evidence for the overall quality of each measurement property were further defined to indicate:

'strong' – consistent findings in multiple studies of good methodological quality OR in one study of excellent quality;

‘moderate’ – consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality;
‘limited’ – one study of fair methodological quality;
‘conflicting’ – conflicting findings; or
‘unknown’ evidence – only studies of poor methodological quality

Where the data entry box is left blank, this signifies no available evidence.

Figure 1: Review of measures used with people with headache – PRISMA flow diagram for article selection (search conducted 1980 to December 2016)