RESEARCH SUBMISSIONS

Development and validation of a novel patient-reported outcome measure in people with episodic migraine and chronic migraine: The Activity Impairment in Migraine Diary

Richard B. Lipton MD[1] | Pranav Gandhi PhD[2] | Jonathan Stokes MBA[2] | Mary Lynn Cala MS[3] |
Christopher J. Evans PhD[3] | Naomi Knoble PhD[4] | Heather L. Gelhorn PhD[4] |
Dennis Revicki PhD[4,5] | Hema N. Viswanathan PhD[6] | David W. Dodick MD[7]

[1]Department of Neurology, Albert Einstein College of Medicine, Bronx, New York, USA

[2]AbbVie, Madison, New Jersey, USA

[3]Endpoint Outcomes, Boston, Massachusetts, USA

[4]Evidera, Bethesda, Maryland, USA

[5]Outcomes Research Consulting, Sarasota, Florida, USA

[6]Allergan, Irvine, California, USA

[7]Department of Neurology, Mayo Clinic, Scottsdale, Arizona, USA

Correspondence
Richard B. Lipton, Albert Einstein College of Medicine, Bronx, NY, USA.
Email: richard.lipton@einsteinmed.org

Abstract

Objective: To evaluate the content validity and psychometric properties of the Activity Impairment in Migraine Diary (AIM-D).

Background: Measuring treatment effects on migraine impairment requires a psychometrically sound patient-reported outcome (PRO) measure developed consistent with U.S. Food and Drug Administration guidance.

Methods: The AIM-D was created from concepts that emerged during qualitative interviews with five clinicians experienced in treating migraine and concept elicitation (CE) interviews with 40 adults with episodic migraine (EM) or chronic migraine (CM). The initial version was refined based on three waves of cognitive interviews with 38 adults with EM or CM and input from a panel of clinical and measurement experts. The AIM-D was psychometrically evaluated using data from 316 adults with EM or CM who participated in a 13-week prospective observational study. Study participants completed PRO assessments including the AIM-D and a daily headache diary. Exploratory and confirmatory factor analysis were used to determine the factor structure. The reliability, validity, and responsiveness of the AIM-D were assessed. Additional PRO measures including the Patient Global Impression – Severity (PGI-S), Migraine Specific Quality of Life Questionnaire, Version 2.1 Role Function-Restrictive domain, and Headache Impact Test were used for psychometric evaluation of the AIM-D.

Results: Based on CE interviews with adults with migraine and input from an expert panel, activity impairment was identified as the target in the preliminary conceptual framework, which had two domains: performance of daily activities (PDAs)

and physical impairment (PI). Revision of the draft AIM-D through multiple rounds of cognitive interviews and expert panel meetings resulted in a content valid 11-item version. Exploratory factor analysis supported both one- and two-domain structures for the AIM-D, which were further supported by confirmatory factor analysis (factor loadings all >0.90). The AIM-D domains (PDA and PI) and total score showed high internal consistency reliability (Cronbach's alpha 0.95–0.97), acceptable test–retest reliability for weekly average scores (intraclass correlation coefficient >0.60 for participants with no change in PGI-S between baseline and week 2), and good convergent and known-groups validity. There was evidence of responsiveness based on changes in PGI-S score and monthly migraine days.

Conclusion: The AIM-D is a content valid and psychometrically sound measure designed to evaluate activity impairment and is suitable for use in clinical trials of preventive treatments for EM or CM.

KEYWORDS

activity impairment, Activity Impairment in Migraine Diary, content validity, migraine, patient-reported outcome, psychometric analysis

INTRODUCTION

Migraine is a sometimes severe and often disabling disease that impairs daily activities and physical functioning.[1] Symptoms can be aggravated by routine physical movement, such as straining and bending over.[2] People with migraine often require bed rest during episodes.[3-5] As a consequence, migraine can interfere with physical functioning, as well as leisure and social activities, and can have a profound impact on emotional and cognitive function.[1,6] Measuring the burden of migraine and the benefits of treatment relies on the use of patient-reported outcome (PRO) measures. In clinical trials of treatments for preventing migraine,[7-10] outcomes in people with episodic migraine (EM) or chronic migraine (CM) have been assessed using various generic and disease-specific PRO measures, including the Migraine Disability Assessment (MIDAS),[11] Headache Impact Test (HIT-6),[12] and Work Productivity and Activity Impairment (WPAI) questionnaire.[13,14] However, although the Food and Drug Administration (FDA) provides guidance on PRO development,[15] at the time this research began, no PRO measures for migraine-related functional impairment were included in labeling for migraine preventive medications in the United States. Subsequently, two PRO measures were included in labeling: the Migraine Specific Quality of Life Questionnaire, Version 2.1 Role Function-Restrictive (MSQ v2.1 RFR) domain and the Migraine Physical Function Impact Diary (MPFID).[16-18] However, the 4-week recall period of the MSQ v2.1 potentially affects the accuracy of the assessments and is subject to recall bias. The MPFID, by contrast, is a daily diary measure, but was still in development and was therefore not available for use.

This paper describes the development and evaluation of the Activity Impairment in Migraine Diary (AIM-D), a new disease-specific PRO measure designed to assess the functional impact of migraine and intended to support labeling for migraine preventive medications. Development of the AIM-D began with a qualitative study to identify the symptom and impact concepts of greatest importance to patients with migraine. Based on findings from this qualitative research, a set of candidate items was developed to assess difficulties in performing daily activities, as well as physical and cognitive impairment due to migraine. Initial development and subsequent refinement of the AIM-D were underpinned by patient input through concept elicitation (CE) and cognitive interviews, input from clinical experts, and FDA recommendations. The psychometric properties of the AIM-D were subsequently evaluated in the context of an observational study in adults with EM or CM. The benefits and limitations of AIM-D as an outcome measure in migraine prevention trials are discussed.

METHODS

Institutional review board approval was obtained for all studies, and written informed consent was obtained from all study participants prior to enrolment. All authors were granted full access to the study data.

Qualitative development of the AIM-D

Qualitative development of the AIM-D is summarized in Figure 1. Consolidated criteria for reporting qualitative research-required information[19] for the qualitative work is provided in Table S1. The AIM-D was developed in accordance with U.S. FDA PRO guidance.[15,20] A targeted literature review was performed to identify existing PRO measures for assessing treatment outcomes in adults with EM or CM.
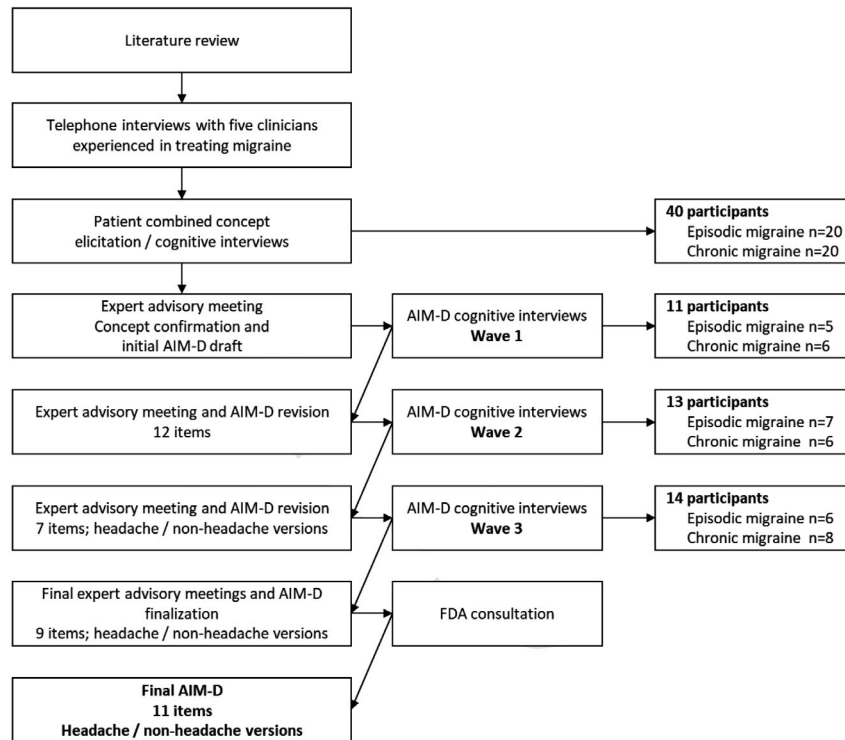
FIGURE 1 Qualitative development of the AIM-D

Concepts relevant for the AIM-D were identified through telephone interviews with five clinicians experienced in treating migraine and face-to-face interviews with adults with EM or CM. The interviews with adults with migraine combined CE and cognitive debriefing of PRO measures, which included draft headache diaries and the Functional Impact of Migraine Questionnaire (FIMQ)[21] but not the AIM-D. Table S2 shows sample questions from the interview guide used for the interviews with adults with migraine. These interviews were audio-recorded, transcribed, anonymized, and coded using ATLAS.ti (Atlas.ti GmbH, Berlin, Germany). The coding scheme, which had been developed based on the target literature review and the objectives of the study, was updated iteratively to reflect the actual terms participants used to describe concepts and to incorporate new concepts that emerged. This process was complemented by clinical guidance (from R.L. and D.D.). Interviews were continued until new concepts ceased to emerge, an outcome known as concept saturation.[22] This was evaluated using saturation grids to determine the adequacy of the sample size and to ensure that no new concepts of interest were likely to be elicited by conducting further interviews. Individual concepts were explored to obtain an in-depth understanding of their meaning by obtaining examples from multiple participants.

In two subsequent in-person advisory meetings with a panel of clinical experts and experts in clinical outcome assessment research, concepts that emerged from these qualitative interviews were used to develop and refine the AIM-D. To minimize recall bias, the AIM-D was developed as a 24-h daily diary. The wording of the draft AIM-D items was informed by the words and phrases patients used

to describe their migraine experience and the concepts of migraine impact. Concepts were selected for inclusion based on their importance and relevance to patients and the extent to which they were aligned with the target measurement concepts. Because the focus of the AIM-D was physical impairment (PI) and performance of daily activities (PDAs), other impacts with a different focus, such as interference with relationships and wanting to be alone, were excluded.

Three additional virtual advisory meetings of the expert panel were conducted to finalize the list of AIM-D items and refine the wording of the items and response options. The second and third of these meetings also aimed to gather clinician feedback on the instrument.

The AIM-D was debriefed in three waves of cognitive interviews with adults with EM or CM. Participants were asked to provide feedback on the instructions, items, and response options and suggest any changes they would make. After each wave of cognitive interviews, participant feedback was considered during one or more virtual meetings of the expert panel. During an additional virtual advisory meeting of the expert panel held after the second wave of interviews, it was decided to create separate versions of the AIM-D for use when headache occurred in the previous 24 h and when it did not.

The preliminary AIM-D conceptual framework included items evaluating PI and difficulties with daily activities. In addition, items assessing cognitive functioning, activity level, and activity limitations were developed outside of the conceptual framework for inclusion in the planned psychometric evaluation, on the advice of the expert panel. Based on the importance of cognitive impacts to patients with

migraine and a review of the qualitative interview studies, the FDA recommended that further consideration be given to evaluating two of the items on cognition (concentrating and thinking clearly) within the AIM-D. Following further discussion with clinical experts, the conceptual framework was modified to include these two items.

Content analysis and psychometric evaluation in a longitudinal observational study

Content analysis and psychometric evaluation of the AIM-D were then assessed in an observational study. The observational study was approved by an institutional review board (Advarra, Columbia, MD) and was conducted in accordance with the Declaration of Helsinki and International Council for Harmonisation E6 guidelines for Good Clinical Practice.

Study design

The observational study was a prospective non-interventional study over 13 weeks (including a 1-week baseline period), conducted at 28 clinical sites in the United States (which included clinical research sites and neurology/pain centers). Participants were treated according to applicable standards of clinical care.

Participants

Enrolment ran from March to May 2019. Participants were recruited either directly by the participating clinical sites or through advertising on social media and then referred to one of the clinical sites for confirmation of eligibility. No formal sample size calculation was performed. However, the aim was to include at least 10 participants per AIM-D item for both the EM and CM subpopulations, in accordance with standard practices.[23]

Eligible participants were English-speaking adults (18–80 years of age) with EM or CM of at least 1 year's duration who met International Classification of Headache Disorders, 3rd edition criteria for migraine with or without aura. The EM group had 4–14 migraine days/month in the previous 3 months, and the CM group had an average of ≥15 headache days/month (with migraine headache on ≥8 days) in the previous 3 months. Participants were recruited who had changed migraine medication or dosing in the 2 weeks prior to enrollment (to facilitate assessment of the AIM-D's responsiveness) or who had been on stable treatment for at least 12 weeks (to facilitate assessment of reliability). People who had changed migraine medication 2–12 weeks previously were excluded.

Potential participants were excluded if they were participating in a clinical trial; had difficulty distinguishing migraine headache from other headache types; had a history of retinal migraine or migraine accompanied by diplopia or decreased consciousness; had responded inadequately to ≥5 prescription preventive medications for migraine; had used opioids or barbiturates for >4 days/month in the previous 3 months (CM only); had a confounding psychiatric condition, a significant risk of self-harm, dementia, epilepsy, or a significant neurological disorder other than migraine; were also suffering from another pain condition; or had a current diagnosis of new persistent daily headache, trigeminal autonomic cephalgia, or painful cranial neuropathy.

AIM-D

Participants completed PRO assessments at home using an eDiary and at the clinical site using an eTablet. They completed the AIM-D daily throughout the study. Each AIM-D item asks respondents to rate level of difficulty on a six-point rating scale ranging from (0) "Not difficult at all" to (5) "I could not do it at all." Three items (errands, leisure outside the home, and strenuous activities) include a response option allowing respondents to indicate that the activity was not planned. The headache and non-headache versions of the AIM-D include the same sets of items and instruct respondents to answer each question based on the level of difficulty experienced "*in the past 24 hours*." However, the headache version instructs respondents to specifically consider the period "*during* [their] *headache*." This is because the impact of migraine on patient functioning on a given day depends on whether or not the patient experiences a headache. This approach also aimed to make the items easier to respond to by anchoring them to a period that is clearly recognizable to respondents. The non-headache version provides a more comprehensive evaluation of functional impairment for days when the respondent does not experience a headache.

In addition to the AIM-D, participants provided daily responses to supplementary items evaluating activity level and activity limitation. Activity level was assessed on a 5-point scale ranging from "No activity – Spent all day lying down" to "Exercised – Brisk walk, running, jogging, biking or other activity for 30 or more minutes," and activity limitation on a 5-point scale ranging from "Not at all limited – I could do everything" to "Extremely limited."

Additional PRO measures for psychometric evaluation

To test the psychometric properties of the AIM-D, participants completed additional PRO assessments using the eDiary or eTablet, including a daily headache diary in which participants recorded whether they had experienced a headache (Yes/No). Other PRO assessments were the EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L),[24,25] FIMQ,[21] HIT-6,[12] Patient-Reported Outcomes Measurement Information System (PROMIS) Pain Interference – Short Form 6a,[26,27] PROMIS Pain Intensity Numeric Rating Scale (NRS),[28] MSQ v2.1,[29,30] and MIDAS.[11] Participants also completed the Patient Global Impression – Severity (PGI-S), a single-item measure that assesses overall severity of migraine symptoms over the previous 7 days on a 5-point scale ranging from "None" to "Very severe"; and

the Patient Global Impression of Change (PGIC), a single-item measure that assesses change in migraine symptoms over time on a 7-point scale ranging from "Very much better" to "Very much worse." The version of the PGIC used in this study asked participants to rate the change in migraine symptoms since the beginning of the study. Further details on these additional PRO assessments, including their timing, are provided in Tables S3 and S4.

Statistical analysis

This was the primary analysis of these data and was based on a prespecified statistical analysis plan for the psychometric analyses. Analyses were conducted using SAS® version 9.4 or later (SAS Institute Inc., Cary, NC). Data for all enrolled participants were used for the content analysis and internal consistency reliability analysis; the full analysis set (FAS), which comprised participants with AIM-D data for baseline and for at least 1 week from day 1 to week 12 (end of study), was used for other analyses. All statistical tests used a two-sided significance level of 0.05. Descriptive statistics were calculated for demographics and baseline characteristics. For the AIM-D, scores were calculated by summing the individual scores (out of 5) for the non-missing items, dividing the result by the number of non-missing items, and multiplying by the total number of items. For exploratory factor analysis and subsequent psychometric evaluation, raw scores were transformed to a 0–100 scale by dividing them by the maximum possible score and multiplying by 100.

The factor structure of the AIM-D was determined using random draws, with one observation per participant. The item-level analyses used data from one headache day and one non-headache day per participant, drawn at random from day 1 through day 28 using the SAS function "ranuni." Floor effects were defined as >30% of participants selecting the minimum response and ceiling effects as >30% of participants selecting the maximum response. Item–item and item–total correlations were calculated as Spearman rank-order correlations using data for the same randomly drawn headache day and non-headache day used to assess floor and ceiling effects.

Exploratory factor analysis was conducted to examine the potential factor structure of the AIM-D using data for each participant from four randomly drawn days, each of which could be a headache or non-headache day. Factors with eigenvalues near to or greater than 1 were favored for retention. Root mean square error of approximation (RMSEA; acceptable if <0.07)[31] and root mean square of residuals were calculated to evaluate model fit. Confirmatory factor analysis was used to confirm the domain structure of the AIM-D using data from three randomly drawn days (headache or non-headache). Weighted least squares (LS) mean and variance-adjusted maximum likelihood estimation was used to estimate the models. Model fit was assessed by calculating comparative fit index (CFI; acceptable if ≥0.9),[32] Tucker–Lewis Index (TLI; acceptable if ≥0.9),[32] RMSEA,[31] and standardized root mean square residual (SRMR; acceptable if <0.08).[33] Factor loadings of ≥0.40 were considered acceptable.[34] For each of the factor analyses, data are presented for one random draw.

Internal consistency reliability was separately assessed for headache and non-headache days by calculating Cronbach's alpha using data for a randomly drawn day (headache or non-headache). A value ≥0.7 is considered good internal consistency reliability.[23,35]

We compared weekly average scores at baseline with weekly average scores at week 2 and monthly average scores at week 4 for participants who selected the same response for the PGI-S at baseline and week 2 or at baseline and week 4, and for patients who indicated "no change" on the PGIC at week 4. A weekly average AIM-D score was calculated if AIM-D scores were available for ≥4 days within a period of seven consecutive days; monthly average AIM-D scores were calculated if AIM-D scores were available for ≥14 days in the relevant 28-day period. In calculating weekly and monthly AIM-D scores, data for all days in the given period were used (i.e., headache and non-headache days were not distinguished from each other). Data for the pairs of time points were compared by a paired t-test. To assess test–retest reliability, intraclass correlation coefficients (ICCs) were calculated using a random-effects analysis of variance model.[36] An ICC of 0.41–0.60 indicates moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 near-perfect agreement.[37]

To examine the convergent validity of the AIM-D, Spearman's rank-order correlations were calculated between AIM-D scores and scores for other PRO measures using baseline data. A correlation coefficient >0.30 among measures of similar concepts indicates moderate convergent validity and a correlation coefficient >0.50 strong convergent validity.[38]

Known-groups validity was evaluated by comparing least squares mean AIM-D scores at baseline between different subgroups of participants, grouped according to number of migraine days, MSQ v2.1 RFR score (dichotomized around the median score: <54 vs. ≥54), and HIT-6 score category.[39] Analysis of covariance was conducted to assess the significance of the differences in AIM-D scores between participant subgroups, with age and sex included in the models as covariates.

Finally, responsiveness was explored by using regression models to evaluate the associations of changes in monthly AIM-D scores between baseline and month 3 with change in PGI-S score (categorized as worsened, no change, improved), percentage change in number of monthly migraine days (no change or increased, reduced by <30%, reduced by ≥30%), change in activity limitation (worsened, no change, improved), and change in HIT-6 total score (no change or increased, reduced by <2.5 points, reduced by ≥2.5 points). Effect sizes for the magnitude of differences between categories were calculated using Cohen's d.[38]

RESULTS

Qualitative development

The targeted literature review identified 17 existing PRO measures that could be used to assess treatment outcomes in adults with EM or CM. Migraine-specific instruments addressing migraine

symptoms, impacts, and/or patient satisfaction and with information available on development of the instrument were subjected to an in-depth review. Evaluation of the seven PRO measures included in the in-depth review revealed limitations in content validity and development according to FDA PRO guidance, and indicated that none of the instruments identified adequately measured the proximal impacts of migraine (Table S5).

Demographics and clinical characteristics of the 40 adults with EM (*n* = 20) or CM (*n* = 20) who participated in the mixed CE/cognitive interviews are shown in Table S6. Concepts that frequently emerged in these interviews included impacts on ability to concentrate (92.5%), ability to move around (80.0%), social activities (80.0%), leisure activities (75.0%), ability to communicate (65.0%), and household activities (65.0%) (Table S7). Saturation of all impact-related concepts was achieved for both EM and CM. The identified impacts were considered during subsequent development of the AIM-D, whose preliminary conceptual framework comprised two domains of items on PI and difficulties with daily activities, as well as a total score.

The AIM-D was debriefed in three waves of cognitive interviews with a total of 38 adults with EM (*n* = 18) or CM (*n* = 20). In addition to the AIM-D items measuring PI and PDAs, items on cognitive functioning ("concentrating," "thinking clearly," and "remembering things"), activity level, and activity limitations were developed and debriefed outside of the AIM-D conceptual framework. Demographics and clinical characteristics of the cognitive interview participants are shown in Table S8. Changes to the initial pool of AIM-D items were made based on expert panel review of participant feedback from the interviews. These changes are summarized in Table S9. To reduce redundancy and shorten the instrument, the three items pertaining to the home were consolidated as two items (household activities at home and leisure activities at home) after the first wave of cognitive interviews (*n* = 11), and daily activities outside the home was removed. Three other items were also removed: moving the head, due to issues in attributing the concept to something other than migraine; moving the body, due to its generic nature and because it was likely captured through another item on walking; and getting around, because participants incorrectly interpreted the item as pertaining to transportation.

Following debriefing of the revised AIM-D in a second wave of cognitive interviews (*n* = 13), moving the head and moving the body were reinstated based on their clinical relevance. In the third wave of cognitive interviews (*n* = 14), participants reported that the instructions, response options, and recall period of the headache and non-headache versions of the AIM-D were easily understood and relevant. Eight participants (57.1%) suggested that relevant concepts such as emotional impact or pain were not covered by the AIM-D, but none of these concepts were reported by more than two participants. Because the instructions, items, and response options were interpreted as intended, no changes to the AIM-D were made based on the results of the third wave of cognitive interviews. However, at the recommendation of the FDA and following further discussion with clinical experts, two of the items on cognition ("concentrating"
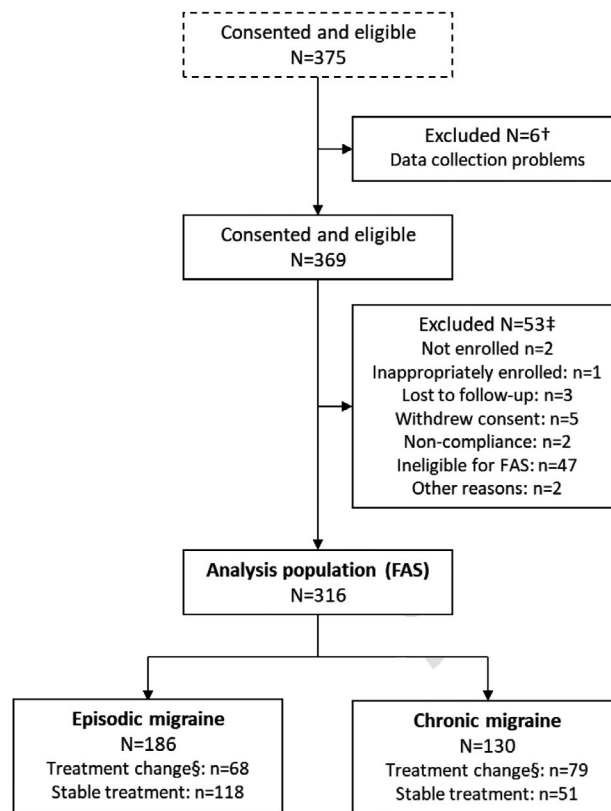
FIGURE 2 Participant disposition. [†]Participants from two clinical sites were excluded because of data collection problems. [‡]Reasons for exclusion were not mutually exclusive. [§]Change of current treatment, including change of dosing (preventive or acute for episodic migraine; preventive for chronic migraine) within the previous 2 weeks. FAS, full analysis set

and "thinking clearly") were added to the preliminary conceptual framework, and the AIM-D was psychometrically evaluated as an 11-item measure (Table S9).

Observational study participants

The observational study was conducted from March to August 2019. A total of 375 participants provided written informed consent and were enrolled, of whom 316 (186 with EM and 130 with CM) were included in the FAS (Figure 2). Forty-seven participants were excluded from the FAS because they did not have AIM-D data for both baseline and at least one subsequent study week.

The mean (standard deviation [SD]) age of participants was 45.0 (12.8) years (range 18–79) and 86.4% of participants were female (Table 1). Most participants were White (74.7%), not Hispanic or Latino (74.1%), and living with a spouse or partner (78.8%). Participants with EM and CM were well balanced in terms of demographics.

The mean time since diagnosis of migraine was 16.5 (13.6) years (range 14 months to 58 years) (Table 1). The most frequently used preventive treatment for migraine was topiramate (26.9% for CM,

13.4% for EM). Two-thirds (66.1%) of participants with EM and 15.4% of participants with CM were not taking a preventive treatment for migraine. Sumatriptan (21.8% overall) was the most frequently used acute treatment for migraine.

Mean (SD) HIT-6 total score at baseline was 62.4 (6.0) (Table 1), indicating a substantial impact on participants' ability to function. Similarly, the mean (SD) PROMIS Pain Interference total score of 60.8 (8.1), which is approximately 1 SD above the US population norm of 50,[40] indicated the impact of pain on participants' lives. For the AIM-D, the mean (SD) total score at baseline was 13.8 (12.8) for participants with EM and 24.3 (18.6) for participants with CM.

Content analysis

Item distributions

On a randomly drawn headache day, item responses were generally well distributed, with only one floor effect (31.8%) for item 6 (walking). Floor effects were observed for all items on a randomly drawn non-headache day and ranged from 78.5% for item 5 (strenuous activities) to 86.0% for item 6 (walking). No ceiling effects were observed.

Item correlations

Item–item correlations on a randomly drawn headache day ranged from 0.65 to 0.93 and item–total correlations ranged from 0.85 to 0.92 (Table S10). On a randomly drawn non-headache day, item–item correlations ranged from 0.61 to 0.89 and item–total correlations ranged from 0.84 to 0.94.

Exploratory factor analysis

One- and two-factor models showed similar fit to the data. All loadings in the single-factor model were >0.40 (Table S11). In the two-factor model, items 1–5, 10, and 11 loaded together (factor loadings 0.420–0.957) and items 6–9 loaded together (0.615–0.894). The correlation between factors 1 and 2 was 0.76.

Confirmatory factor analysis

Confirmatory factor analysis supported one- and two-factor structures (Table 2). Factor loadings for all items were >0.90. Model fit was acceptable for both the one-factor structure (CFI = 0.997, TLI = 0.996) and the two-factor structure (CFI = 0.999, TLI = 0.998). For the one-factor structure, RMSEA was 0.14 and SRMR was 0.016; RMSEA was 0.10 and SRMR 0.010 for the two-factor structure. Based on these findings, and in accordance with both the AIM-D conceptual framework and recommendations from the FDA in their

assessment of the qualitative research, both the one- and two-factor models were evaluated in subsequent psychometric analyses.

Final scoring of the AIM-D

The final version of the AIM-D comprises a PDAs domain (items 1–5, 10, and 11) and a PI domain (items 6–9). A total score is also calculated (Figure 3). Domain scores and AIM-D total score are calculated on a 0–100 scale, with a higher score indicating a greater level of impairment. The minimum number of non-missing responses for a score to be calculated is 4 out of 7 for the PDA domain, 2 out of 4 for the PI domain, and 6 out of 11 for AIM-D total score. Otherwise, the score is set to missing. Across day 1 through day 28 of the observational study, data for the AIM-D were missing for an average of 5.3 days (18.9%).

Psychometric evaluation

Internal consistency reliability

Internal consistency reliability was high. For the total sample (all participants, EM or CM), Cronbach's alpha on a randomly drawn headache day was 0.97 for the PDA domain, 0.95 for the PI domain, and 0.98 for AIM-D total score (Table S12). Deletion of individual items had little impact on Cronbach's alpha. Similar results were obtained for participants with EM and participants with CM (Table S12), and for a randomly drawn non-headache day (Table S13).

Test–retest reliability

Reproducibility of AIM-D scores was good. For the total sample, ICCs for AIM-D domain scores and total score were >0.60 for participants with no change in PGI-S between baseline and week 2 (Table 3). ICCs for EM and CM were also >0.60. For participants with no change between baseline and week 4 in PGI-S (Table S14) or PGIC (Table S15), ICCs for the total sample, EM, and CM were similarly all >0.60.

Convergent validity

The AIM-D demonstrated construct validity. For the total sample, AIM-D domain scores and total score at baseline showed moderate correlations with activity level (−0.41 to −0.45), moderate to strong correlations with number of headache days (0.49–0.58), and strong correlations with activity limitation (0.80–0.86) and number of migraine days (0.59–0.69) (Table 4). Correlations with PGI-S (0.53–0.55), PROMIS Pain Interference total score (0.54–0.57), and FIMQ total score (0.56–0.60) were strong and those with HIT-6 total score (0.36–0.38) and MSQ domain scores (−0.36 to −0.50) were

TABLE 1 Demographics and baseline characteristics

	Total sample (N = 316)	Episodic migraine (N = 186)	Chronic migraine (N = 130)
Demographics			
Age, years			
Mean (SD)	45.0 (12.8)	43.9 (12.8)	46.7 (12.7)
Range	18–79	18–79	21–75
Sex, n (%)			
Female	273 (86.4)	160 (86.0)	113 (86.9)
Race, n (%)[a]			
American Indian/Alaskan Native	4 (1.3)	3 (1.6)	1 (0.8)
Asian	22 (7.0)	14 (7.5)	8 (6.2)
Black or African American	58 (18.4)	33 (17.7)	25 (19.2)
Native Hawaiian/Pacific Islander	0	0	0
White	236 (74.7)	141 (75.8)	95 (73.1)
Other	5 (1.6)	2 (1.1)	3 (2.3)
Ethnicity, n (%)			
Hispanic or Latino	82 (25.9)	54 (29.0)	28 (21.5)
Not Hispanic or Latino	234 (74.1)	132 (71.0)	102 (78.5)
Living/domestic situation, n (%)			
Living alone	64 (20.3)	40 (21.5)	24 (18.5)
Living with spouse or partner	249 (78.8)	145 (78.0)	104 (80.0)
Other	3 (0.9)	1 (0.5)	2 (1.5)
Clinical characteristics[b]			
Diagnosis, n (%)			
Migraine without aura	241 (76.3)	143 (76.9)	98 (75.4)
Migraine with aura	143 (45.3)	81 (43.5)	62 (47.7)
Years since diagnosis of migraine			
Mean (SD)	16.5 (13.6)	15.7 (13.0)	17.7 (14.4)
Range	1.2–58.0	1.2–55.0	1.3–58.0
History of 4–14 migraine days/month, n (%)		186 (100.0)	0
Average number of migraine days/month,[c] n (%)			
4–5 days		67 (36.0)	
6–7 days		51 (27.4)	
8–9 days		26 (14.0)	
10–14 days		42 (22.6)	
History of 15 or more headache days/month, n (%)		0	130 (100.0)
Preventive medication for migraine ever prescribed, n (%)			
None	143 (45.3)	123 (66.1)	20 (15.4)
Baseline PRO assessments[d]			
HIT-6 total score			
Mean (SD)	62.4 (6.0)	61.7 (6.0)	63.4 (5.9)
PROMIS Pain Interference total score			
Mean (SD)	60.8 (8.1)	59.7 (8.8)	62.4 (6.9)
PROMIS Pain Intensity NRS			
Mean (SD)	5.9 (2.3)	5.6 (2.4)	6.2 (2.0)
MSQ v2.1			
Role Function-Restrictive			
Mean (SD)	51.7 (23.6)	55.0 (23.6)	47.1 (23.0)

TABLE 1 (Continued)

	Total sample (N = 316)	Episodic migraine (N = 186)	Chronic migraine (N = 130)
Role Function-Preventive			
Mean (SD)	64.4 (24.6)	66.2 (24.0)	61.8 (25.2)
Emotional Function			
Mean (SD)	62.2 (30.2)	65.8 (29.6)	57.1 (30.5)
MIDAS			
Mean (SD)	71.8 (85.4)	66.6 (91.7)	79.4 (75.0)
Missing, *n*	5	2	3
EQ-5D-5L			
Utility score			
Mean (SD)	0.86 (0.11)	0.88 (0.10)	0.82 (0.11)
Missing, *n*	48	28	20
VAS			
Mean (SD)	79.2 (13.6)	81.7 (12.0)	75.6 (15.0)
Missing, *n*	48	28	20
FIMQ			
Mean (SD)	46.6 (19.1)	43.0 (18.1)	51.6 (19.6)
AIM-D[e]			
PDA domain			
Mean (SD)	20.5 (17.7)	15.7 (14.0)	27.3 (20.1)
PI domain			
Mean (SD)	14.5 (15.0)	11.0 (11.7)	19.5 (17.5)
AIM-D total score			
Mean (SD)	18.1 (16.3)	13.8 (12.8)	24.3 (18.6)

Abbreviations: AIM-D, Activity Impairment in Migraine Diary; EQ-5D-5L, EuroQoL 5 Dimensions 5 Levels; FIMQ, Functional Impact of Migraine Questionnaire; HIT-6, Headache Impact Test; MIDAS, Migraine Disability Assessment; MSQ v2.1, Migraine Specific Quality of Life Questionnaire, Version 2.1; NRS, Numeric Rating Scale; PDA, performance of daily activities; PI, physical impairment; PRO, patient-reported outcome; PROMIS, Patient-Reported Outcomes Measurement Information System; SD, standard deviation; VAS, Visual Analog Scale.

[a]Not mutually exclusive.

[b]Reported by the clinical site.

[c]Only collected for participants with episodic migraine.

[d]No assessments were missing, except where indicated.

[e]AIM-D scores were transformed to a 0–100 scale.

moderate. Similar results were obtained for participants with EM (Table S16) and participants with CM (Table S17).

Known-groups validity

The AIM-D also demonstrated known-groups validity. For EM, mean AIM-D domain scores and total score at baseline were higher for participants with an average of 10–14 migraine days/month than for those with 4–5 migraine days/month (Table 5). Similarly, for CM, mean AIM-D scores at baseline were higher for participants with an average of 14–22 or 23–28 migraine days/month than for those with 0–7 migraine days/month (Table 6). Known-groups validity was also demonstrated when baseline AIM-D scores were analyzed according to MSQ RFR score (Table S18) and

HIT-6 total score (minimal/mild impact vs. moderate/severe impact) (Table S19).

Responsiveness

The AIM-D showed evidence of being responsive to changes in migraine frequency and severity. For the total sample and for participants with EM, changes in AIM-D domain scores and total score between baseline and month 3 were significantly higher in participants with a ≥1-point improvement in PGI-S score than in those whose PGI-S score worsened by ≥1 points (Table 7). Changes in AIM-D domain score and total score between baseline and month 3 were also higher in participants with a ≥30% decrease in number of monthly migraine days compared to those with no change or an

TABLE 2 Confirmatory factor analysis: Randomly drawn headache or non-headache day (day 1 through day 28), total sample

	One-factor model	Two-factor model	
	Factor 1	Factor 1 (PDA)	Factor 2 (PI)
Factor loadings for individual items			
1. Chores	0.973	0.977	
2. Errands	0.990	0.991	
3. Leisure home	0.970	0.975	
4. Leisure outside	0.975	0.977	
5. Strenuous activities	0.951	0.955	
6. Walk	0.958		0.968
7. Move body	0.974		0.983
8. Bend forward	0.962		0.976
9. Move head	0.954		0.967
10. Concentrate	0.958	0.963	
11. Think clearly	0.958	0.964	
Factor intercorrelation			
Factor 1 with factor 2		0.957	
Model fit			
χ^2 (df)*	352.13 (44)	191.55 (43)	
p-value	<0.01	<0.01	
CFI	0.997	0.999	
TLI	0.996	0.998	
RMSEA (90% CI)	0.14 (0.12–0.15)	0.10 (0.08–0.11)	
SRMR	0.016	0.010	

* indicates *p* less than or equal to 0.05.

Abbreviations: CFI, comparative fit index; CI, confidence interval; df, degrees of freedom; PDAs, performance of daily activities; PI, physical impairment; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker–Lewis Index.
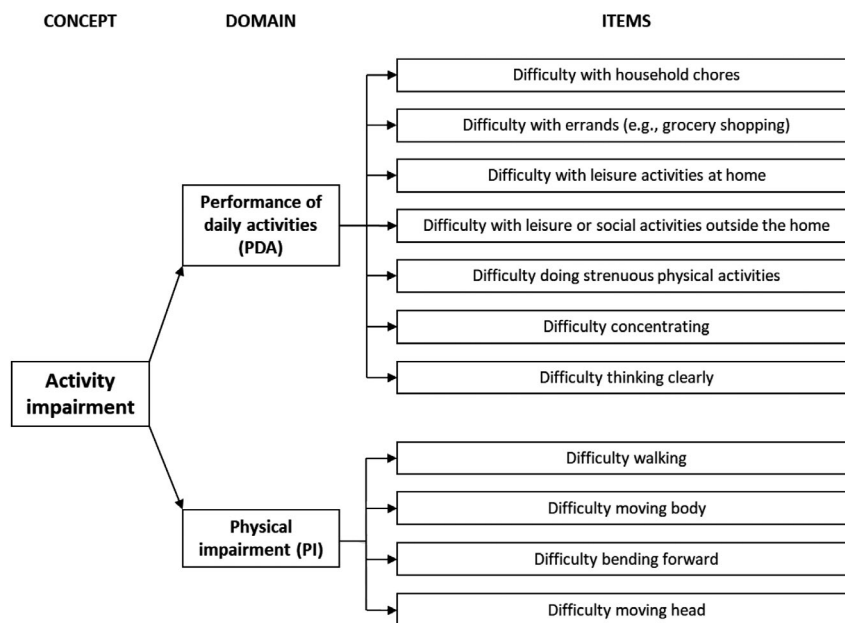

FIGURE 3 Final conceptual framework for the AIM-D

TABLE 3 Test–retest reliability: Participants with no change in PGI-S between baseline and week 2

| Characteristic | N | Weekly average score, mean (SD) | | Difference | t^a | p-value[a] | ICC |
		Baseline	Week 2				
Total sample							
PDA domain	147	19.3 (16.3)	17.2 (15.0)	−2.1	−2.24	0.027	0.73
PI domain	147	13.2 (14.0)	13.0 (13.6)	−0.2	−0.19	0.849	0.74
AIM-D total score	147	16.9 (15.0)	15.6 (14.0)	−1.4	−1.56	0.120	0.73
Episodic migraine							
PDA domain	87	15.3 (13.7)	13.9 (12.0)	−1.4	−1.21	0.228	0.67
PI domain	87	10.3 (11.5)	10.2 (10.8)	−0.1	−0.16	0.874	0.74
AIM-D total score	87	13.3 (12.5)	12.4 (11.1)	−0.9	−0.90	0.372	0.69
Chronic migraine							
PDA domain	60	25.1 (18.0)	21.9 (17.6)	−3.2	−1.95	0.056	0.74
PI domain	60	17.4 (16.2)	17.2 (16.2)	−0.2	−0.12	0.908	0.72
AIM-D total score	60	22.1 (16.8)	20.1 (16.5)	−2.0	−1.29	0.201	0.73

Abbreviations: AIM-D, Activity Impairment in Migraine Diary; ICC, intraclass correlation coefficient; PDAs, performance of daily activities; PGI-S, Patient Global Impression – Severity; PI, physical impairment; SD, standard deviation.
[a]Paired t-test.

TABLE 4 Convergent validity: Correlations between AIM-D scores and other measures at baseline, total sample

	PDA domain	PI domain	AIM-D total score
Patient questionnaire			
Activity level	−0.45***	−0.41***	−0.44***
Activity limitation	0.86***	0.80***	0.86***
Daily headache diary			
Number of headache days	0.58***	0.49***	0.56***
Number of migraine days	0.69***	0.59***	0.67***
HIT−6 total score	0.37***	0.36***	0.38***
MSQ v2.1			
Role Function-Restrictive	−0.44***	−0.43***	−0.45***
Role Function-Preventive	−0.49***	−0.49***	−0.50***
Emotional Function	−0.38***	−0.36***	−0.38***
PGI-S	0.54***	0.53***	0.55***
PROMIS Pain Interference total score	0.56***	0.54***	0.57***
PROMIS Pain Intensity NRS	0.41***	0.40***	0.42***
FIMQ total score	0.59***	0.56***	0.60***

Note: Spearman rank-order correlations.
Abbreviations: AIM-D, Activity Impairment in Migraine Diary; FIMQ, Functional Impact of Migraine Questionnaire; HIT-6, Headache Impact Test; MSQ v2.1, Migraine Specific Quality of Life Questionnaire, Version 2.1; NRS, Numeric Rating Scale; PDA, performance of daily activities; PGI-S, Patient Global Impression – Severity; PI, physical impairment; PROMIS, Patient-Reported Outcomes Measurement Information System.
***$p < 0.0001$.

increase in the number of migraine days (Table S20), and in those whose activity limitation improved versus those whose activity limitation did not change or worsened (Table S21). AIM-D total score and domain scores were not found to differ significantly according to change in HIT-6 score between baseline and month 3 (Table S22).

DISCUSSION

PRO measures developed in accordance with FDA guidance are important for capturing the patient perspective on treatment efficacy in clinical trials. The AIM-D was developed and evaluated following FDA guidance[15,41,42] as a measure for assessing activity impairment, with items and response options that are relevant for patients with EM or CM. The development and evaluation process included CE (where saturation of concepts was demonstrated); patient feedback on each version of the measure; regular input from clinical experts; and the assessment of quantitative evidence. Our qualitative work indicated that activity impairment was a crucial aspect of the patient experience of migraine. In developing the AIM-D, we focused on capturing PDAs and PI. Factor analysis supported one- and two-factor structures for the AIM-D. This provided a basis for calculating PDA and PI domain scores and AIM-D total score. Internal consistency reliability was excellent and reproducibility was good. The AIM-D domains demonstrated construct validity and known-groups validity, and showed evidence of being responsive to changes in symptom severity and frequency, and in activity limitation, but not to changes in HIT-6 score.

Other PRO measures developed per FDA guidance are now available. The 13-item MPFID was published after work on the AIM-D had begun, and is included in the label for erenumab-aooe.[43] As with the AIM-D, it evaluates the physical impact of migraine, with a 24-h recall period that gives precision and limits the influence of recall bias.[16,17] The item content of the MPFID is similar to that of

TABLE 5 Known-groups validity at baseline: Average number of month migraine days, episodic migraine

		Average number of migraine days									F	p-value	p-value for pairwise comparisons[a]
		4–5		6–7		8–9		10–14					
	N (total)	N	LS mean (SE)	N	LS mean (SE)	N	LS mean (SE)	N	LS mean (SE)				
PDA domain	128	37	9.7 (1.8)	15	13.9 (2.8)	37	16.3 (1.8)	39	20.8 (1.8)	4.03	0.002	1 vs. 2: 0.663 1 vs. 3: 0.089 **1 vs. 4: <0.001** 2 vs. 3: 0.919 2 vs. 4: 0.243 3 vs. 4: 0.373
PI domain	128	37	6.1 (1.7)	15	9.8 (2.7)	37	11.6 (1.7)	39	15.8 (1.7)	3.53	0.005	1 vs. 2: 0.719 1 vs. 3: 0.176 **1 vs. 4: 0.002** 2 vs. 3: 0.960 2 vs. 4: 0.322 3 vs. 4: 0.385
AIM-D total score	128	37	8.3 (1.7)	15	12.4 (2.7)	37	14.2 (1.7)	39	18.8 (1.7)	3.93	0.002	1 vs. 2: 0.653 1 vs. 3: 0.120 **1 vs. 4: <0.001** 2 vs. 3: 0.953 2 vs. 4: 0.260 3 vs. 4: 0.317

Bolded values correspond to p less than or equal to 0.05.

Abbreviations: AIM-D, Activity Impairment in Migraine Diary; LS, least squares; SE, standard error.

[a]Scheffe's test.

TABLE 6 Known-groups validity at baseline: Number of migraine days, chronic migraine

| | | Number of migraine days | | | | | | | | F | p-value | p-value for pairwise comparisons[a] |
| | | 0–7 | | 8–13 | | 14–22 | | 23–28 | | | | |
	N (total)	N	LS mean (SE)	N	LS mean (SE)	N	LS mean (SE)	N	LS mean (SE)			
PDA domain	130	31	13.6 (3.2)	32	21.9 (3.2)	32	34.8 (3.2)	35	37.5 (3.0)	8.15	<0.0001	1 vs. 2: 0.333 **1 vs. 3: <0.001** **1 vs. 4: <0.0001** **2 vs. 3: 0.049** **2 vs. 4: 0.007** 3 vs. 4: 0.945
PI domain	130	31	10.1 (2.9)	32	15.2 (2.9)	32	26.3 (2.9)	35	25.7 (2.7)	5.99	<0.0001	1 vs. 2: 0.665 **1 vs. 3: 0.002** **1 vs. 4: 0.002** 2 vs. 3: 0.064 2 vs. 4: 0.075 3 vs. 4: 0.999
AIM-D total score	130	31	12.2 (3.0)	32	19.3 (3.0)	32	31.4 (3.0)	35	33.0 (2.8)	7.71	<0.0001	1 vs. 2: 0.425 **1 vs. 3: <0.001** **1 vs. 4: <0.0001** **2 vs. 3: 0.045** **2 vs. 4: 0.012** 3 vs. 4: 0.985

Bolded values correspond to p less than or equal to 0.05.

Abbreviations: AIM-D, Activity Impairment in Migraine Diary; LS, least squares; PDA, performance of daily activities; PI, physical impairment; SE, standard error.

[a]Scheffe's test.

TABLE 7 Responsiveness between baseline and month 3 based on change in PGI-S

AIM-D score change	Change in PGI-S						F^b	p-value	Effect size
	1: Worsened by ≥1 points		2: No change[a]		3: Improved by ≥1 points				
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)			
Total sample									
PDA domain	47	−0.2 (11.3)	119	−2.4 (12.2)	86	−8.3 (13.9)	8.01	<0.001	1 vs. 2: −0.18
									1 vs. 3: −0.62
									2 vs. 3: −0.46
PI domain	47	2.3 (10.4)	119	−0.1 (10.5)	86	−4.6 (12.5)	6.88	0.001	1 vs. 2: −0.23
									1 vs. 3: −0.59
									2 vs. 3: −0.40
AIM-D total score	47	0.7 (10.6)	119	−1.5 (11.2)	86	−6.8 (13.2)	7.91	0.001	1 vs. 2: −0.20
									1 vs. 3: −0.61
									2 vs. 3: −0.45
Episodic migraine									
PDA domain	34	−0.6 (10.4)	71	−2.9 (10.4)	44	−8.8 (14.2)	5.45	0.005	1 vs. 2: −0.22
									1 vs. 3: −0.64
									2 vs. 3: −0.49
PI domain	34	2.0 (9.6)	71	−0.6 (9.0)	44	−4.9 (11.5)	4.95	0.008	1 vs. 2: −0.29
									1 vs. 3: −0.65
									2 vs. 3: −0.43
AIM-D total score	34	0.4 (9.9)	71	−2.0 (9.4)	44	−7.3 (13.2)	5.51	0.005	1 vs. 2: −0.25
									1 vs. 3: −0.65
									2 vs. 3: −0.48
Chronic migraine									
PDA domain	13	0.8 (13.6)	48	−1.6 (14.6)	42	−7.8 (13.7)	2.94	0.057	1 vs. 2: −0.16
									1 vs. 3: −0.63
									2 vs. 3: −0.44
PI domain	13	3.2 (12.9)	48	0.8 (12.6)	42	−4.3 (13.6)	2.48	0.089	1 vs. 2: −0.19
									1 vs. 3: −0.56
									2 vs. 3: −0.39
AIM-D total score	13	1.8 (12.9)	48	−0.6 (13.5)	42	−6.4 (13.4)	2.86	0.062	1 vs. 2: −0.18
									1 vs. 3: −0.61
									2 vs. 3: −0.42

Abbreviations: AIM-D, Activity Impairment in Migraine Diary; PDA, performance of daily activities; PGI-S, Patient Global Impression – Severity; PI, physical impairment; SD, standard deviation.

[a]Same response selected at baseline and month 3.

[b]General linear model.

the AIM-D, which further supports the content validity of the AIM-D. However, despite their overlap, the AIM-D and MPFID measure different aspects of the migraine burden.

One of the most widely used disease-specific measures in migraine is the MSQ v2.1, which evaluates how frequently migraine limits and prevents work and social and activities, as well as the emotional impact of migraine.[29,30] The MSQ v2.1 RFR domain is included in the label for galcanezumab-gnlm.[44] While its longer 4-week recall period potentially makes it difficult for respondents

to accurately average the impacts of migraine over time, it may focus patients on what is most salient. However, the MSQ v2.1 does not distinguish between headache and non-headache days. Moreover, the AIM-D captures meaningful physical and cognitive impacts not captured by the MSQ v2.1, such as difficulty walking and difficulty thinking clearly, and assesses severity rather than frequency of impacts. Considering also the convergent validity between AIM-D scores and scores for the MSQ v2.1 domains (correlations −0.36 to −0.50) and for the HIT-6 (0.36–0.38), together

with the fact that three of the HIT-6 items have a 4-week recall period, the AIM-D may be a useful complement to PRO measures with longer recall periods.

An important limitation of the present analysis is the observational, non-interventional design of the validation study. As is often seen with this type of study design, changes in disease status were modest, despite design aspects aimed at producing observable changes including active recruitment of participants who had recently changed treatment. For this reason, the responsiveness analysis should be interpreted with caution and should be replicated in datasets that are better suited to this kind of analysis, such as data from randomized controlled trials. The smaller size of the CM sample compared with the EM sample may have limited the power of certain analyses for CM. In addition, item-level analyses based on randomly drawn days and psychometric analyses based on weekly or monthly averages had the potential to be affected by the relative frequencies of headache and non-headache days. However, the contrasting item distributions for headache and non-headache days do not imply lack of reliability; rather, they suggest sensitivity to state-dependent change. The eligibility criteria applied are similar to those typically used in clinical trials of treatments for migraine prevention, which supports the generalizability of the findings. Strengths of the AIM-D include the availability of headache and non-headache versions, which enables day-to-day variability in the impact of migraine to be captured. By instructing respondents to focus on the period "*during* [their] *headache*," the headache version specifically captures their experience during headache (rather than during the full 24-h recall period, where impairment may be pronounced during migraine and more modest at other times). The non-headache version meets the growing emphasis on assessing interictal non-headache symptoms and impacts relating to migraine. Finally, although the AIM-D was completed daily, the average monthly rate of missing data was low (5.3 out of 28 days).

CONCLUSION

The AIM-D is a content valid and psychometrically sound measure of activity impairment with migraine. It was developed in alignment with FDA PRO guidance, using qualitative data from substantial numbers of adults with EM or CM, and its measurement properties were confirmed in a longitudinal observational study. The robust quantitative data support evaluation of the PDA and PI domains and AIM-D total score as endpoints in clinical trials of preventive treatments in patients with EM or CM. In addition to providing summary measures of burden of illness and benefit of treatment, the AIM-D will facilitate comparisons of migraine burden on headache, premonitory, postdromal, and interictal days. Future research is needed to further evaluate clinically meaningful changes for the AIM-D domains. The potential for using the AIM-D to evaluate patient outcomes in routine clinical practice, and in digital applications for capturing real-world data, should also be investigated.

CONFLICT OF INTEREST

RBL receives research support from the NIH and the National Headache Foundation; holds stock options in Biohaven Holdings and Ctrl M Health; serves as a consultant, advisory board member, or has received honoraria from: Abbvie (Allergan), Amgen, Axsome, Biohaven, Biovision, Dr. Reddy's (Promius), Electrocore, Eli Lilly, Equinox, GlaxoSmithKline, Grifols, Lundbeck (Alder), Merck, Pernix, Pfizer, Satsuma, and Teva; and receives royalties from Wolff's Headache 7th and 8th Editions (Oxford Press University, Wiley, and Informa). PG and JS are employees of AbbVie and may own AbbVie stock. MLC was, during this work, an employee of Endpoint Outcomes, which received funds from Allergan to carry out the research. CJE has received support for research and consulting services from Allergan (at the time the research was conducted) and AbbVie (currently). NK was an employee of Evidera at the time of this research and received funding from Allergan for time spent conducting this research. HLG is an employee of Evidera, a company that received funding from Allergan for time spent conducting this research. DR has received support for research and consulting services from Allergan, AbbVie, and Amgen. HNV was an employee and shareholder of Allergan (now AbbVie) at the time the study was conducted. DWD reports the following conflicts of interest within the past 12 months: consulting: AEON, Amgen, Clexio, Cerecin, Cooltech, Ctrl M, Allergan, Alder, Biohaven, GSK, Linpharma, Lundbeck, Promius, Eli Lilly, eNeura, Novartis, Impel, Satsuma, Theranica, WL Gore, Nocira, XoC, Zosano, Upjohn (Division of Pfizer), Pieris, Praxis, Revance, Equinox. Honoraria: CME Outfitters, Curry Rockefeller Group, DeepBench, Global Access Meetings, KLJ Associates, Academy for Continued Healthcare Learning, Majallin LLC, Medlogix Communications, MJH Lifesciences, Miller Medical Communications, Southern Headache Society (MAHEC), WebMD Health/Medscape, Wolters Kluwer, Oxford University Press, Cambridge University Press. Research Support: Department of Defense, National Institutes of Health, Henry Jackson Foundation, Sperling Foundation,

AUTHOR CONTRIBUTIONS

Study concept and design: Richard B. Lipton, Pranav Gandhi, Jonathan Stokes, Christopher J. Evans, Naomi Knoble, Heather L. Gelhorn, Dennis Revicki, Hema N. Viswanathan. *Acquisition of data*: Richard B. Lipton, Mary Lynn Cala, Christopher J. Evans, Naomi Knoble, Heather L. Gelhorn, Dennis Revicki. *Analysis and interpretation of data*: Pranav Gandhi, Jonathan Stokes, Mary Lynn Cala, Christopher J. Evans, Naomi Knoble, Heather L. Gelhorn, Dennis Revicki, Hema N. Viswanathan, David W. Dodick. *Drafting of the manuscript*: Richard B. Lipton, Pranav Gandhi, Jonathan Stokes, Mary Lynn Cala, Naomi Knoble, Hema N. Viswanathan. *Revising it for intellectual content*: Richard B. Lipton, Pranav Gandhi, Jonathan Stokes, Mary Lynn Cala, Christopher J. Evans, Naomi Knoble, Heather L. Gelhorn, Dennis Revicki, Hema N. Viswanathan, David W. Dodick. *Final approval of the completed manuscript*: Richard B. Lipton, Pranav Gandhi, Jonathan Stokes, Mary Lynn Cala, Christopher J. Evans, Naomi Knoble, Heather L. Gelhorn, Dennis Revicki, Hema N. Viswanathan, David W. Dodick.

ORCID

David W. Dodick https://orcid.org/0000-0002-9486-6790

REFERENCES

1. Mannix S, Skalicky A, Buse DC, et al. Measuring the impact of migraine for evaluating outcomes of preventive treatments for migraine headaches. *Health Qual Life Outcomes*. 2016;14:143.
2. Spierings EL, Ranke AH, Honkoop PC. Precipitating and aggravating factors of migraine versus tension-type headache. *Headache*. 2001;41:554-558.
3. Brandes JL. Global trends in migraine care: results from the MAZE survey. *CNS Drugs*. 2002;16(Suppl 1):13-18.
4. Hu XH, Markson LE, Lipton RB, Stewart WF, Berger ML. Burden of migraine in the United States: disability and economic costs. *Arch Intern Med*. 1999;159:813-818.
5. Agosti R. Migraine burden of disease: from the patient's experience to a socio-economic view. *Headache*. 2018;58(Suppl 1):17-32.
6. Buse DC, Rupnow MF, Lipton RB. Assessing and managing all aspects of migraine: migraine attacks, migraine-related functional impairment, common comorbidities, and quality of life. *Mayo Clin Proc*. 2009;84:422-435.
7. Peng KP, Wang SJ. Migraine diagnosis: screening items, instruments, and scales. *Acta Anaesthesiol Taiwan*. 2012;50:69-73.
8. Buse DC, Lipton RB, Hallström Y, et al. Migraine-related disability, impact, and health-related quality of life among patients with episodic migraine receiving preventive treatment with erenumab. *Cephalalgia*. 2018;38:1622-1631.
9. Skljarevski V, Matharu M, Millen BA, Ossipov MH, Kim BK, Yang JY. Efficacy and safety of galcanezumab for the prevention of episodic migraine: results of the EVOLVE-2 Phase 3 randomized controlled clinical trial. *Cephalalgia*. 2018;38:1442-1454.
10. Lipton RB, Cohen JM, Gandhi SK, Yang R, Yeung PP, Buse DC. Effect of fremanezumab on quality of life and productivity in patients with chronic migraine. *Neurology*. 2020;95:e878-e888.
11. Stewart WF, Lipton RB, Kolodner K, Liberman J, Sawyer J. Reliability of the migraine disability assessment score in a population-based sample of headache sufferers. *Cephalalgia*. 1999;19(2):107-114; discussion 174.
12. 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;12:963-974.
13. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4:353-365.
14. McGinley JS, Houts CR, Nishida TK, et al. Systematic review of outcomes and endpoints in preventive migraine clinical trials. *Headache*. 2021;61:253-262.
15. FDA. *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. U.S. Department of Health and Human Services Food and Drug Administration; 2009.
16. Hareendran A, Mannix S, Skalicky A, et al. Development and exploration of the content validity of a patient-reported outcome measure to evaluate the impact of migraine- the migraine physical function impact diary (MPFID). *Health Qual Life Outcomes*. 2017;15:224.
17. Kawata AK, Hsieh R, Bender R, et al. Psychometric evaluation of a novel instrument assessing the impact of migraine on physical functioning: the migraine physical function impact diary. *Headache*. 2017;57:1385-1398.
18. Speck RM, Shalhoub H, Wyrwich KW, et al. Psychometric validation of the role function restrictive domain of the Migraine Specific Quality-of-Life Questionnaire Version 2.1 electronic patient-reported outcome in patients with episodic and chronic migraine. *Headache*. 2019;59:756-774.
19. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19:349-357.
20. Cala ML, Graham CA, Lipton RB, et al. The Activity Impairment in Migraine Diary (AIM-D): a novel migraine-specific patient-reported outcome measure to assess functioning based on activity impairment related to migraine in episodic and chronic migraine patients. Poster presented at the American Headache Society 60th Annual Scientific Meeting, San Francisco, CA, 2018.
21. Rosa KR, Lai H, Yang M. Evaluation of the Assessment of Chronic Migraine Impacts (ACM-I) and Assessment of Chronic Migraine Symptoms (ACM-S). Poster presented at the American Headache Society 56th Annual Scientific Meeting, Los Angeles, CA, 2014.
22. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? *Qual Health Res*. 2017;27:591-608.
23. Nunnally JC, Bernstein IH. *Psychometric Theory*. McGraw-Hill; 1994.
24. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727-1736.
25. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22:1717-1727.
26. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain*. 2010;150:173-182.
27. Askew RL, Cook KF, Revicki DA, Cella D, Amtmann D. Evidence from diverse clinical populations supported clinical validity of

PROMIS pain interference and pain behavior. *J Clin Epidemiol.* 2016;73:103-111.

28. Revicki DA, Chen W-H, Harnam N, et al. Development and psychometric analysis of the PROMIS pain behavior item bank. *Pain.* 2009;146:158-169.

29. Jhingran P, Davis SM, LaVange LM, Miller DW, Helms RW. MSQ: Migraine-Specific Quality-of-Life Questionnaire. Further investigation of the factor structure. *Pharmacoeconomics.* 1998;13:707-717.

30. 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;40:204-215.

31. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. *Testing Structural Equation Models.* Sage Publications; 1993:136-162.

32. Hu L, Bentler PM. Evaluating model fit. In: Hoyle RH, ed. *Structural Equation Modeling: Concepts, Issues, and Applications.* Sage Publications; 1995:76-99.

33. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1-55.

34. Muthén LK, Muthén BO. *Mplus User's Guide.* Muthén & Muthén; 1998–2010.

35. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297-334.

36. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials.* 1991;12:142S-158S.

37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

38. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Lawrence Erlbaum Associates; 1988.

39. Bayliss M, Batenhorst A. *The HIT-6™: A User's Guide.* QualityMetric, Inc.; 2002.

40. PROMIS. *PROMIS Pain Interference Scoring Manual;* 2020.

41. FDA. *Discussion Document for Patient-Focused Drug Development. Public Workshop on Guidance 2: Methods to Identify What is Important to Patients;* 2018.

42. FDA. *Discussion Document for Patient-Focused Drug Development. Public Workshop on Guidance 3: Select, Develop or Modify Fit-for-Purpose Clinical Outcome Assessments.* 2018.

43. *AIMOVIG™ (Erenumab-Aooe) Injection, for Subcutaneous Use.* Full prescribing information; 2018.

44. *EMGALITY™ (Galcanezumab-Gnlm) Injection, for Subcutaneous Use.* Full prescribing information; 2018.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.