Psychometric Analysis Plan

Study title: An observational cohort study to understand the psychometric properties of the developing MiCOAS migraine-specific patient-reported outcome measures

Short title: MICOAS QUANT1

Study purpose: Data is collected to allow for psychometric analyses to be completed for the developing MiCOAS patient-reported outcome measures to understand their dimensionality, reliability, and the validity of inferences made from the scores

Study phase: NA

Sponsor Protocol:

Version Date: 03/13/2026

Version #: 1.3

CONFIDENTIAL

VECTOR PSYCHOMETRIC GROUP, LLC
CHAPEL HILL, NORTH CAROLINA U.S.A.

Carrie R. Houts, PhD
Director, Psychometrics
crhouts@vpgcentral.com

This page has been intentionally left blank.

VERSION HISTORY

Date	Updated By	Version	Description
11/12/2024	CRH	1.1	Reorganization to separate recall-based and DA analyses for clarity; update DA analyses to examine weekly, biweekly, and monthly summary scores
02/07/2025	CRH	1.2	Update DIF analyses to remove gender and race as grouping variables
03/13/2026	SL	1.3	Update references

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Abbreviation	Definition
ANOVA	Analysis of variance
BAEM	Bock-Aitkin expectation maximization
CFA	Confirmatory factor analysis
CIFA	Confirmatory item factor analysis
CM	Chronic migraine
CTT	Classical test theory
DA	Daily assessment
DIF	Differential item functioning
eCDF	Empirical cumulative distribution function
ePDF	Empirical probability distribution function
EFA	Exploratory factor analysis
EIFA	Exploratory item factor analysis
EF	Emotional functioning
EM	Episodic migraine
FDA	Food and Drug Administration
HIT-6	Headache Impact Test
ICC	Intraclass correlation coefficient
IFA	Item factor analysis
IRT	Item response theory
M	Mean
MFIQ	Migraine Functional Impact Questionnaire
MiCOAS	Migraine Clinical Outcome Assessment System
MiCOAS DA	MiCOAS daily assessment
MiCOAS 7d	MiCOAS 7-day recall based measures
MiCOAS 14d	MiCOAS 7-day recall based measures
MMD	Monthly migraine day
MOS-Cog-R	Medical Outcome Study – Cognitive Function - Revised
MSQ	Migraine-specific Quality of Life Questionnaire
PAP	Psychometric analysis plan
PF	Physical functioning
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Status
PRO	Patient reported outcome
RMSEA	Root mean square error of approximation
SCF	Subjective cognitive functioning
SD	Standard deviation
SE	Standard error
SF	Social role functioning
SEM	Standard error of measurement
SMA	Subjective mental acuity

SoA	Schedule of assessments
TLI	Tucker-Lewis Index
VPG	Vector Psychometric Group, LLC
wABC	Weighted area between the curve

INTRODUCTION

The Migraine Clinical Outcome Assessment System (MiCOAS) is partnership between Vector Psychometric Group, LLC (VPG) and Albert Einstein Medical School (co-primary investigators: RJ Wirth, PhD and Richard B. Lipton, MD) that has received a U.S. Food and Drug Administration (FDA) grant to develop a core set of outcomes for use in migraine clinical trials. Over the last 4 years, the MiCOAS research group has completed several rounds of concept elicitation interviews with people living with migraine to understand what is important to them in terms of symptomology, impact, function, and outcomes and identify gaps in the existing patient-reported outcome (PRO) measures. More recently, draft instruments have been developed based on information provided by people living with migraine via qualitative interviews.

The MiCOAS research group is currently developing an observational, short-term longitudinal study (QUANT1) that will allow for the initial investigation of the psychometric properties of the MiCOAS measures being developed: a daily assessment (DA), and two recall-based assessments, one with a 7-day (7d) and the other with a 14-day (14d) recall period. The 21-item DA is intended to create 3 measures (symptoms [9 items], subjective mental acuity [SMA; 5 items], and physical functioning [PF; 8 items]) with one additional general item. The recall-based questionnaires are intended to create 4 measures (PF [9 items], subjective cognitive functioning [SCF; 10 items], social role functioning [SF; 7 items], and emotional functioning [EF; 6 items]), along with 4 additional general items. The purpose of the current document is to detail analyses that are planned for completion to thoroughly examine these measures and begin to accumulate evidence that is appropriate for regulatory decision making to determine if these PRO measures are fit-for-purpose in migraine trials.

STUDY DESIGN

The full study design of the MiCOAS is available in the full study protocol (2023-001 MiCOAS Protocol Approved 20231218.pdf), but we provide an overview of the salient design features here. The QUANT1 study will be observational/non-interventional in nature and will collect data longitudinally over the course of 8 weeks/2 months. In total, the MiCOAS research team sought to have a sample size of approximately 1000 respondents. Given the developing MiCOAS measures under investigation, as well as the need to include reference measures that allowed for validation analyses of the scores coming from the MICOAS DA and recall-based measures, a cohort design was selected in an effort to reduce respondent burden within any given assessment time. In short, 3 cohorts were designed, the first (approximate n = 150) has a focus on collecting information used to examine the MiCOAS DA over an 8-week period, the second (approximate n = 350) has a focus on the MiCOAS DA and the 14-day recall based MiCOAS measures over a 4-week period, and the third (approximate n = 500) is focused on MiCOAS DA and the 7-day recall based MiCOAS measures over a 4-week period.

The following psychometric analysis plan (PAP) documents analyses to assess the psychometric properties of the MiCOAS PRO measures and investigate the validity of the resulting summary scores from these measures in a convenience sample selected, in part, to mirror demographic characteristics and other key features of

migraine randomized controlled trials for either preventive or acute migraine medications or medical devices. Methods and software to be used will be described in sufficient detail to allow replication.

Below, we describe scales/measures and relevant timepoints that will be used in the current analyses.

MEASURES

- There are three MiCOAS recall-based measures totaling 32 items:
  - Physical function (PF; 9 items)
  - Subjective cognitive function (SCF; 10 items)
  - Social role function (SF; 7 items).
  - Emotional function (EF; 6 items)

  Additionally, 4 general function items ("On how many days did you have migraine symptoms?", "How often did you feel good?", "How often were you able to enjoy all or part of the day?", and "How often did you want to spend time alone in a comfortable environment?") are also included. Two versions of the recall-based measures have been created, one with a 7-day recall period and the other with a 14-day recall period (MiCOAS 7d and MiCOAS 14d, respectively). The physical function and social/emotional role function items generally use a 5-category ordinal response scale with frequency labels (i.e., 0 = Never to 4 = Always), while the cognitive function items tend to use 5-category ordinal responses with verbal labels on a difficulty scale (i.e., 0 = Not difficult at all to 4 = Unable to do). While the scoring algorithm will be determined based on the results of this study, it is expected all MiCOAS scores will be obtained such that higher values are associated with less positive outcomes (e.g., worse physical function).

  The MiCOAS 14d questionnaire is administered at Baseline, Week 2, Week 4, Week 6, and Week 8 in Cohort 1. In Cohort 2, the MiCOAS 14d questionnaire is administered at Baseline, Week 2, and Week 4. In Cohort 3, the MiCOAS 7d questionnaire is administered at Baseline, Week1, Week 2, and Week 4.

- The MiCOAS DA (MiCOAS DA) is a 23-item assessment intended to be completed by participants on a daily basis. The content of the 23 items is spread across one general item and 3 individual measures:
  - Symptoms (9 items)
  - Subjective mental acuity (SMA; 5 items)
  - PF (8 items)

  As noted, the MiCOAS DA also includes one general item asking about the participants day with respect to their migraine experience. Two items ("Did you have any migraine symptoms?" and "Did you have a good day?") of the MiCOAS DA use a Yes/No response scale and the remaining 21 items use 5-category ordinal response scales, with verbal labels varying depending on item content. Most symptom items use a severity response scale (0 = Did not have to 4 = Severe verbal labels) while most of the mental

acuity and function items use a difficulty response scale (verbal labels ranging from 0= Not at all difficult to 4 = Unable to do). As such, when the scoring algorithm is finalized, it is expected that higher scores will be associated with worse outcomes (e.g., worse physical functioning, more severe symptomology).

Additionally, due to the desire to explore definitions of monthly migraine days (MMDs), a common outcome in preventive migraine trials, several other symptom and symptom-related items not included in the MiCOAS DA based on relevance/importance to patients, such as duration and unilaterality of head pain, are also included in the daily assessment to allow for the definition of a migraine day and, when summarized, MMDs.

In Cohort 1 (DA emphasis), the daily assessments, including the MiCOAS DA, will be completed on a daily basis for 8 weeks (2 months). In Cohorts 2 and 3, participants will be asked to complete the assessment on a daily basis for 4 weeks (1 month).

- The Migraine Functional Impact Questionnaire (MFIQ v2; Kawata et al., 2019) is a 26-item questionnaire with a 7-day recall period intended to create measures of:
    - Physical function (5 items)
    - Usual activities (11 items)
    - Social function (5 items)
    - Emotional function (5 items)

  All items use an ordinal, 5-category response scale, with verbal labels varying based on item content. Higher scores on each of the MFIQ domains are associated with worse outcomes (e.g., poorer physical function).

  In all three cohorts, the MFIQ v2 is administered at Baseline, Week 2, and Week 4. Additionally, in Cohort 1 the MFIQ v2 will also be administered at Week 6 and Week 8.

- The Migraine-specific Quality of Life Questionnaire (MSQ v2.1; e.g., Martin et al., 2000) is a 14-item measure with a 4-week recall period intended to assess 3 domains:
    - Role function - restrictive (7 items)
    - Role function – preventive (4 items)
    - Emotional function (3 items)

  All items use a 6-category ordered response scale with verbal labels ranging from 1 = None of the time to 6 = All of the time. For all three domains of the MSQ v2.1, summed scores from the items contributing to each domain are calculated and then rescaled to a 0 to 100 metric. For all domains, higher scores are associated with better quality of life/outcomes (e.g., fewer role restrictions due to migraine).

In all three cohorts, the MSQ v2.1 will be administered at Baseline and Week 4/Month 1. Additionally, in Cohort 1 the MSQ v2.1 will be administered at Week 8/Month 2

- The 6-item Headache Impact Test (HIT-6; Bayliss & Batenhorst, 2002) is designed to measure the adverse impacts that migraine and headaches have on respondents' lives over the last 4 weeks. All 6 items use a 5-category ordinal response scale with verbal labels ranging from Never to Always. The HIT-6 total score is found as a summed score. The weighting of response options as they contribute to this sum score is as follows: Never (6 points), Rarely (8 points), Sometimes (10 points), Very Often (11 points), and Always (13 points). Rather than using more typical 1 to 5 response-category values, this weighting scheme was empirically determined in the initial publication of the HIT-6 (Kosinski et al., 2003) to maximize the degree to which the HIT-6 summed score mirrored scores derived using IRT estimation and more items from the adaptive administration of the HIT item bank; scores range from 36 to 79 with higher scores indicate greater impact due to migraine/headaches. Thresholds for severity categories have been suggested for the HIT-6 scores (Bayliss & Batenhorst, 2002) which are: 49 or less = Little to no impact; 50-55 = Moderate impact; 56-59 = Substantial impact; and 60 or greater = Severe impact.

  In all three cohorts, the HIT-6 is administered at Baseline and Week 4. Additionally, in Cohort 1 the HIT-6 is administered at Week 8.

- The revised MOS Cognitive Function measure (MOS-Cog-R; Yarlas et al., 2013) is a 6-item assessment intended to assess cognitive function in adults. The MOS-Cog-R version used for the current work will have a 7-day recall period. The item content of the measure covers such topics as concentration and thinking, confusion, attention, and subjective reaction time. All items of the MOS-Cog-R use a 5-category response scale, with verbal labels ranging from "No time" to "All of the time." A total score is calculated by summing the 6 individual item responses and higher total score values are associated with better cognitive functioning.

  In all three cohorts, the 7-day recall MOS-COG-R will be administered weekly (Cohort 1: Baseline and Week 1 through Week 8; Cohorts 2 and 3: Baseline and Week 1 through Week 4).

- A patient global impression of status (PGIS) item, asking participants their impression of the severity of their overall status over the last 7 or 14 days (depending on cohort).

  In Cohort 1, the PGIS will be administered with a 14-day recall period at Baseline, Week 2, Week 4 Week 6, and Week 8. In Cohort 2, the PGIS will be administered with a 14-day recall period at Baseline, Week 2, and Week 4. In Cohort 3, the PGIS is planned for administration with a 7-day recall period at Baseline, Week 1, Week 2, Week 3, and Week 4.

- A patient global impression of change (PGIC) item, asking participants their impression of how their overall status has changed "since the start of the study".

  In Cohort 1, the PGIC will be administered at Week 2, Week 4, Week 6, and Week 8. In Cohort 2, the PGIC will be administered at Week 2 and Week 4. In Cohort 3, the PGIC will be administered at Week 1, Week 2, Week 3, and Week 4.

For ease of review for timing of assessments, a schedule of assessments (SoA) is provided in Figure 1.

Figure 1. SoA for the MiCOAS QUANT1 study

Measure	Baseline Day 0	Month 1 Week 1 Day 7	Month 1 Week 2 Day 14	Month 1 Week 3 Day 21	Month 1 Week 4 Day 28	Month 2 Week 5 Day 35	Month 2 Week 6 Day 42	Month 2 Week 7 Day 49	Month 2 Week 8 Day 56
Cohort 1 - DA Emphasis (n=150)									
MiCOAS									
MiCOAS DA		■	■	■	■	■	■	■	■
MiCOAS 7d									
MiCOAS 14d	■		■		■		■		■
Cognition									
MOS-Cog-R 7d									
Headache PROM									
HIT-6 (4-week recall)	■				■				■
MFIQ v2 (7-day recall)	■		■		■		■		■
MSQ v2.1 (4-week recall)	■				■				■
General PROM									
PGIS (14-day recall)	■				■		■		■
PGIC					■		■		■
Cohort 2 - 14-day PROM Emphasis (n=350)									
MiCOAS						■	■	■	■
MiCOAS DA		■	■	■	■	■	■	■	■
MiCOAS 7d						■	■	■	■
MiCOAS 14d	■		■		■	■	■	■	■
Cognition						■	■	■	■
MOS-Cog-R 7d						■	■	■	■
Headache PROM						■	■	■	■
HIT-6 (4-week recall)	■				■	■	■	■	■
MFIQ v2 (7-day recall)	■		■		■	■	■	■	■
MSQ v2.1 (4-week recall)	■				■	■	■	■	■
General PROM						■	■	■	■
PGIS (14-day recall)	■		■		■	■	■	■	■
PGIC			■		■	■	■	■	■
Cohort 3 - 7-day PROM Emphasis (n=500)									
MiCOAS						■	■	■	■
MiCOAS DA		■	■	■	■	■	■	■	■
MiCOAS 7d	■	■	■	■	■	■	■	■	■
MiCOAS 14d						■	■	■	■
Cognition						■	■	■	■
MOS-Cog-R 7d	■	■	■	■	■	■	■	■	■
Headache PROM						■	■	■	■
HIT-6 (4-week recall)	■				■	■	■	■	■
MFIQ v2 (7-day recall)	■	■	■	■	■	■	■	■	■
MSQ v2.1 (4-week recall)	■				■	■	■	■	■
General PROM						■	■	■	■
PGIS (7-day recall)	■	■	■	■	■	■	■	■	■
PGIC		■	■	■	■	■	■	■	■

Note. MiCOAS DD = MiCOAS Daily Diary. MiCOAS 7d = MiCOAS 7-day recall based questionnaire. MiCOAS 14d = MiCOAS 14-day recall-based questionnaire. MOS-Cog-R-7d = Medical Outcome Study- Cognition measure revised with 7-day recall. MOS-Cog-R-14d = Medical Outcome Study- Cognition measure revised with 14-day recall. HIT-6 = 6-item HEadache Impact Test short form. MFIQ v2 = Migraine Functional Impact Questionnaire version 2.0. MSQ v2.1 = Migraine-specific Quality of Life Questionnaire version 2.1. PGIS = Patient global impression of status. PGIC = Patient global impression of change.

GENERAL METHODS

Analyses designed to investigate the psychometric properties of the individual items and measures of the developing MiCOAS migraine PRO questionnaires (DA and recall-based) in an adult sample are detailed below. Analyses will include assessments of test-retest reliability, as well as several different analyses intended to provide evidence regarding the validity of the scores from the MiCOAS PRO measures in people living with migraine, including convergent and discriminant correlations, known groups analyses, sensitivity to change analyses, and preliminary analyses to investigate meaningful within-person change threshold (i.e., a cut-point for the minimum amount of change in each target PRO measure score necessary for an individual to experience clinically meaningful improvement) for each measure score obtained from the questionnaire. Table shells are provided to detail the expected presentation of results; the final structure of tables may be modified from the shells to optimize presentation and interpretability of observed results. In addition to the specified analyses, appropriate visualizations to support the analyses (e.g., histograms of MiCOAS item response distributions, scatter plots of correlations, box plots to visualize known-group analyses) will be generated and provided in a supplemental appendix to the psychometric report.

DATA HANDLING AND IMPUTATIONS

Prior to the primary analyses (specified below), the individual observation's response patterns will be reviewed to attempt to identify and remove inattentive responders, such as those that provide the same numeric response to all items within a given assessment window or those who complete a large number of items in an extremely limited amount of time (indicating they were likely not reading or providing thoughtful responses to items). Specifically, response patterns will be examined for "straight line" responders, where all or most items receive the same response, including items that are of opposite valence from one another (e.g., hypothetical items of "I had a great day" and "I had a terrible day" both receiving responses of Strongly Agree would be considered straight lining) and the item responses for that assessment will be examined for potential removal prior to analysis.

Classical Test Theory (CTT) analyses require complete item-level data (within an individual assessment). Observations with missing MiCOAS measure item responses at the CTT-target timepoint will be excluded from CTT analyses.

Item factor analysis (IFA) and item response theory (IRT) analyses do not require data on all candidate items to produce scores; therefore, no imputations or exclusions are planned in preparing the data for IFA/IRT analyses.

For monthly MiCOAS DA scores, it is required that 15 or more of the DA days are completed out of the possible 28 days in each month.

MMD is the number of "migraine days" over the selected 28-day periods, as operationally defined below. It will be calculated when at least 15 days of the DA are completed out of each 28-day period (Day 1 through

Day 28 for Month 1, Day 29 through Day 56 for Month 2). If the number of missing days in the respective 28-day period is > 13 days, MMD will be coded as missing. When the number of missing days in the respective 28-day period is between 1 and 13, a normalization procedure will be applied by multiplying the proportion of migraine days (number of migraine days divided by total number of non-missing days) by 28. For example, if a subject reports 5 migraine days over a total of 20 non-missing DA days, the normalized MMD value would be 7 (i.e., (5/20)*28 = 7).

The definition of a migraine day will be adapted from the International Headache Society International Classification of Headache Disorders-3 criteria of migraine. Specifically, a qualified migraine day is defined as any calendar day with a headache (head pain of mild, moderate, or severe intensity), with or without aura, lasting for ≥ 30 minutes, and meeting both of the following criteria (A and B):

A. ≥ 2 of the following headache characteristics:
    a. Unilateral location
    b. Pulsating quality (throbbing),
    c. Moderate or severe pain intensity,
    d. Aggravation by or causing avoidance of routine physical activity (e.g., walking to climbing stairs)
B. ≥ 1 of the following:
    a. Nausea and/or vomiting,
    b. Photophobia and phonophobia

DESCRIPTIVE STATISTICS

To provide an understanding of the sample and context for reported validation analyses, descriptive statistics (e.g., means, standard deviations [SDs], medians, response frequency distributions) for the baseline demographic and descriptor variables (e.g., age, gender, race, ethnicity) listed in the shell for Table 1 will be summarized to describe the sample characteristics and allow for broad comparison to other populations (typical migraine clinical trial population, broader community of people living with migraine, etc.).

Table 1. Baseline summary of categorical demographic and descriptor variables

Variable	Value	
Age (years)		
Mean (SD)	xx.xx(x.x)	
Median		
Q1, Q3		
Min, max		
	Percent	n
Sex		
Female	xx.x	xxx
Male	xx.x	xxx
Other	xx.x	xxx
Prefer not to answer	xx.x	xxx
Race		
American Indian or Alaska Native	xx.x	xxx
Asian	xx.x	xxx
Black or African American	xx.x	xxx
White	xx.x	xxx
Other	xx.x	xxx
Prefer not to answer	xx.x	xxx
Ethnicity		
Hispanic, Latino, or Spanish origin	xx.x	xxx
Not Hispanic	xx.x	xxx
Prefer not to answer		
EM	xx.x	xxx
4-9 MHD	xx.x	xxx
10-14 MHD	xx.x	xxx
CM	xx.x	xxx
15-21 MHD	xx.x	xxx
22-27 MHD	xx.x	xxx

METHODS FOR RECALL-BASED MICOAS MEASURES

Summary descriptive values for relevant continuous reference measures (e.g., MMDs, MSQ V2.1, MFIQ V2.0, HIT-6) will be reported at target timepoints (e.g., Table 2) and response distributions will be reported for categorical reference variables (i.e., PGIS and PGIC item responses) at target timepoints (e.g., Table 3) For the MiCOAS 7d version, target timepoints are Baseline (Day 0), Week 1 (Day 7), Week 2 (Day 14), Week 3 (Day 21), and Week 4 (Day 28). For the MiCOAS 14d versions, target timepoints are Baseline (Day 0), Week 2 (Day 14), Week 4 (Day 28), Week 6 (Day 42), and Week 8 (Day 56) and will use data from both Cohort 1 and 2.

Table 2. Descriptive statistics for continuous reference measures in Cohort 3 that will be used to examine the MiCOAS 7d scores

	Baseline (Day 0)			Week 1 (Day 7)			Week 2 (Day 14)			Week 3 (Day 21)			Week 4 (Day 28)		
Measure	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Self-reported number of migraine days	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Cognition															
MOS-Cog-R 7d	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Headache PRO measures															
HIT-6 (4-week recall)	xxx	xx.x	xx.x										xxx	xx.x	xx.x
MSQ v2.1 (4-week recall)															
Role Function - restrictive	xxx	xx.x	xx.x										xxx	xx.x	xx.x
Role Function - preventative	xxx	xx.x	xx.x										xxx	xx.x	xx.x
Emotional function	xxx	xx.x	xx.x										xxx	xx.x	xx.x
MFIQ v2 (7-day recall)															
Physical Function	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Usual Activities	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Social Function	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Emotional Function	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x	xxx	xx.x	xx.x
Clinical measure															
MMD													xxx	xx.x	xx.x

Similar table created for continuous reference variables in Cohort 1 and 2 at target timepoints of Baseline (Day 0), Week 2 (Day 14), Week 4 (Day 28), Week 6 (Day 42), and Week 8 (Day 56), that will used to examine the MiCOAS 14d scores

Table 3. Descriptive statistics and response distributions for categorical reference variables in Cohort 3 that will be used to examine the MiCOAS 7d scores

Measure	Item content	Cat0	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6
		Baseline (Day 0)						
HIT-6 item 1	How often pain is severe?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
HIT-6 item 6	How often felt too tired to work?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
MSQ item 5	How often migraine limit ability to concentrate?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
MFIQ item 14	How difficult...activities required concentration?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
PGIS	Severity of overall status over the past week	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
		Week 1 (Day 7)						
MFIQ item 14	How difficult...activities required concentration?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
PGIS	Severity of overall status over the past week	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
PGIC	Since the start of the study, my overall status is:	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
		Week 2 (Day 14)						
MFIQ item 14	How difficult...activities required concentration?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
PGIS	Severity of overall status over the past week	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
PGIC	Since the start of the study, my overall status is:	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
		Week 3 (Day 21)						
MFIQ item 14	How difficult...activities required concentration?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
PGIS	Severity of overall status over the past week	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
PGIC	Since the start of the study, my overall status is:	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
		Week 4 (Day 28)						
HIT-6 item 1	How often pain is severe?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
HIT-6 item 6	How often felt too tired to work?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
MSQ item 5	How often migraine limit ability to concentrate?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
MFIQ item 14	How difficult...activities required concentration?	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
PGIS	Severity of overall status over the past week	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---
PGIC	Since the start of the study, my overall status is:	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)

Note. For HIT-6 items, Cat0 = Never, Cat1 = Rarely, Cat2 = Sometimes, Cat3 = Very often, Cat4 = Always. For MSQ item 5, Cat0 = NOne of the time, Cat1 = A little bit of the time, Cat2 = Some of the time, Cat3 = A good bit of the time, Cat4 = Most of the time, Cat5 = All of the time. For MFIQ item 14, Cat0 = Not difficult, Cat1 = A little difficult, Cat2 = Moderately difficult, Cat4 = Very difficult, Cat5= Extremely difficult. For PGIS, Cat0 = None, Cat1 = Mild, Cat2 = Moderate, Cat3 = Severe, Cat4 = Very Severe. For PGIC, Cat0 = Very much worse, Cat1 = Much worse, Cat2 = Minimally worse, Cat3 = No change, Cat4 = Minimally improved, Cat5 = Much improved, and Cat6 = Very much improved

Similar table created for categorical reference variables in Cohort 1 and 2 at target timepoints of Baseline (Day 0), Week 2 (Day 14), Week 4 (Day 28), Week 6 (Day 42), and Week 8 (Day 56), that will used to examine the MiCOAS 14d scores

For clarity, the item content and response scale of each of the *a priori* recall-based measures (which are identical across the 7d and 14d recall periods) is reported below in Table 4 to Table 7.

Table 4. Items included in the MiCOAS 7d and 14d PF measure

Item content	Response scale
Was it difficult for you to do your usual day-to-day activities?	Difficulty
How often were you too tired to do your regular daily activities?	Frequency
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	Frequency
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	Difficulty
Was it difficult for you to do your regular household chores?	Difficulty
Was it difficult for you to do activities because you were sensitive to light?	Difficulty
Was it difficult for you to do activities because you were sensitive to smells?	Difficulty
Was it difficult for you to do activities because you were sensitive to sound?	Difficulty
Was it difficult for you to do errands?	Difficulty

Note. Difficulty response scale is Not difficult at all, A little difficult, Somewhat difficult, Very difficult, Unable to do. Frequency response scale is Never, Rarely, Sometimes, Very often, Always.

Table 5. Items included in the MiCOAS 7d and 14d SCF measure

Item content	Response scale
Was it difficult for you to concentrate?	Difficulty
Was it difficult for you to speak clearly?	Difficulty
Was it difficult for you to remember the right word for something?	Difficulty
Was it difficult for you to have a conversation?	Difficulty
Was it difficult for you to understand what was said to you?	Difficulty
Was it difficult for you to understand what you were reading?	Difficulty
Was it difficult for you to make day-to-day decisions?	Difficulty
Was it difficult for you to carry out tasks that required many steps?	Difficulty
Was it difficult for you to remember things that were important to you?	Difficulty
Was it difficult for you to do tasks that required you to concentrate?	Difficulty

Note. Difficulty response scale is Not difficult at all, A little difficult, Somewhat difficult, Very difficult, Unable to do

Table 6. Items included in the MiCOAS 7d and 14d SF measure

Item content	Response scale
Was it difficult for you to take care of people or pets that you usually take care of?	Difficulty
Was it difficult for you to manage your usual tasks and responsibilities at school or at work?	Difficulty
Was it difficult for you to keep plans you made?	Difficulty
How often were you reluctant to make plans with other people?	Frequency
How often did you have to cancel or change your plans for social or recreation activities?	Frequency
Was it difficult for you to enjoy social activity with other people?	Difficulty
Was it difficult for you to take part in activities you do for fun?	Difficulty

Note. Difficulty response scale is Not difficult at all, A little difficult, Somewhat difficult, Very difficult, Unable to do. Frequency response scale is Never, Rarely, Sometimes, Very often, Always.

Table 7. Items included in the MiCOAS 7d and 14d EF measure

Item content	Response scale
How often did you feel you did not have control over your life?	Frequency
How often did you feel worried about having a migraine attack?	Frequency
How often were you concerned that your migraine attacks would affect other people's lives?	Frequency
How often did you feel sad or depressed because of migraine?	Frequency
How often did you feel frustrated because of migraine?	Frequency
How often did you feel irritable because of migraine?	Frequency

Note. Frequency response scale is Never, Rarely, Sometimes, Very often, Always.

ITEM-LEVEL DESCRIPTIVE STATISTICS

For the MiCOAS 7d and 14d measures, item summaries will be provided at the previously defined target timepoints (e.g., Table 8). Item-level summaries will be examined for floor effect, ceiling effect, and missing data to identify items that are performing sub-optimally in the migraine patient sample (i.e., if a large number of observed responses occur in the least severe [floor effect] or most severe [ceiling effect] response category for a given item, this may indicate that item is not informative/well-calibrated to the sample). Rather than specifying a certain percentage to define floor and ceiling effects (which should change based on the number of response categories for each item), floor and ceiling effects are defined when the mode of the observed response option falls in the lowest response category (floor effect) or the highest/response category (ceiling effect).

PSYCHOMETRIC ANALYSES

Analyses to investigate the psychometric properties of the MiCOAS 7d and 14d items will be completed.

After reviewing the raw data response characteristics, the response distributions of each 7d and 14d item in the data to be submitted to psychometric analysis will be reviewed and sparseness, when present, will be noted. If there are limited (fewer than 10) observed responses on a given response category, this category will be collapsed into the next lowest response category (e.g., if there are only 3 responses in the 4: "Severe" response option for a given MiCOAS measure item, those responses would be collapsed/recoded into the 3: "Moderate" response category) prior to further analysis. This collapsing/recoding is done to ensure that there are sufficient observations within each analyzed response category to allow for stable and precise parameter estimates to be obtained for each item in the psychometric analyses.

As an initial tool to investigate how the items of each of the various MiCOAS measures are related to one another within each questionnaire, polychoric correlations among the (collapsed, as needed) items of the MiCOAS 7d and 14d measures will be reported (Table 9). For the items that are included in the item sets but not assigned to any *a priori* measure (4 from the recall-based measures), polychoric correlation values will be reviewed to determine which, if any, measure these items could be assigned to.

Table 8. Item descriptives and observed frequencies of the MiCOAS 7d version items at Baseline (Day 0)

Item content	N	M	Mode	SD	Cat0	Cat1	Cat2	Cat3	Cat4	Does not apply
How often did you feel good?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often were you able to enjoy all or part of a day?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you want to spend time alone in a comfortable environment?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do your usual day-to-day activities?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often were you too tired to do your regular daily activities?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do your regular household chores?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do activities because you were sensitive to light?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do activities because you were sensitive to smells?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do activities because you were sensitive to sound?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do errands?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to concentrate?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to speak clearly?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to remember the right word for something?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to have a conversation?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to understand what was said to you?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to understand what you were reading?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to make day-to-day decisions?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to carry out tasks that required many steps?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to remember things that were important to you?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to do tasks that required you to concentrate?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to take care of people or pets that you usually take care of?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
Was it difficult for you to manage your usual tasks and responsibilities at school or at work?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
Was it difficult for you to keep plans you made?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often were you reluctant to make plans with other people?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you have to cancel or change your plans for social or recreation activities?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to enjoy social activity with other people?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
Was it difficult for you to take part in activities you do for fun?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you feel you did not have control over your life?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you feel worried about having a migraine attack?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often were you concerned that your migraine attacks would affect other people's lives?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you feel sad or depressed because of migraine?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you feel frustrated because of migraine?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---
How often did you feel irritable because of migraine?	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---

Similar tables provided for 7d item at other target timepoints and for all MiCOAS 14d items at all target timepoints

Table 9. Polychoric correlations among items of the MiCOAS 7d version

Item content	Item #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
How many days did you have migraine symptoms?	1	1																																			
How often did you feel good?	2	0.xx	1																																		
How often were you able to enjoy all or part of a day?	3	0.xx	0.xx	1																																	
How often did you want to spend time alone in a comfortable environment?	4	0.xx	0.xx	0.xx	1																																
Was it difficult for you to do your usual day-to-day activities?	5	0.xx	0.xx	0.xx	0.xx	1																															
How often were you too tired to do your regular daily activities?	6	0.xx	0.xx	0.xx	0.xx	0.xx	1																														
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	7	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																													
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	8	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																												
Was it difficult for you to do your regular household chores?	9	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																											
Was it difficult for you to do activities because you were sensitive to light?	10	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																										
Was it difficult for you to do activities because you were sensitive to smells?	11	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																									
Was it difficult for you to do activities because you were sensitive to sound?	12	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																								
Was it difficult for you to do errands?	13	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																							
Was it difficult for you to concentrate?	14	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																						
Was it difficult for you to speak clearly?	15	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																					
Was it difficult for you to remember the right word for something?	16	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																				
Was it difficult for you to have a conversation?	17	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																			
Was it difficult for you to understand what was said to you?	18	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																		
Was it difficult for you to understand what you were reading?	19	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																	
Was it difficult for you to make day-to-day decisions?	20	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1																
Was it difficult for you to carry out tasks that required many steps?	21	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1															
Was it difficult for you to remember things that were important to you?	22	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1														
Was it difficult for you to do tasks that required you to concentrate?	23	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1													
Was it difficult for you to take care of people or pets that you usually take care of?	24	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1												
Was it difficult for you to manage your usual tasks and responsibilities at school or at work?	25	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1											
Was it difficult for you to keep plans you made?	26	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1										
How often were you reluctant to make plans with other people?	27	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1									
How often did you have to cancel or change your plans for social or recreation activities?	28	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1								
Was it difficult for you to enjoy social activity with other people?	29	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1							
Was it difficult for you to take part in activities you do for fun?	30	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1						
How often did you feel you did not have control over your life?	31	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1					
How often did you feel worried about having a migraine attack?	32	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1				
How often were you concerned that your migraine attacks would affect other people's lives?	33	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1			
How often did you feel sad or depressed because of migraine?	34	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1		
How often did you feel frustrated because of migraine?	35	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1	
How often did you feel irritable because of migraine?	36	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1

For ease of review, polychoric correlation tables will also be provided for each item subset contributing to an *a priori* measure and for the MiCOAS 14d versions

Factor Analytic Modeling

CIFA models will also be examined for the MiCOAS 7d and 14d questionnaires, separately, using Baseline data. Similar models and processes will be used for the MiCOAS 7d and 14d, so general methods are described but note that the analyses below will be performed on the 7-day items and then again on the 14-day items. While it is expected that the dimensionality results will be similar across the MiCOAS 7d and 14d, there is no *a priori* constraint that the same final model must be derived from these analyses.

For the MiCOAS 7- and 14-day recall measure items, which have limited response categories making continuous data analysis methods unwise (e.g., Wirth & Edwards, 2007), item factor analytic modeling will be conducted in flexMIRT 3.7 (Cai, 2024). In flexMIRT, a full-information categorical estimation method, specifically Bock-Aitkin expectation-maximization (BAEM) estimation (Bock & Aitkin, 1981), will be employed. SEs of item parameters will be estimated using the supplemented expectation maximization algorithm (Cai, 2008). If high-dimensional models (i.e., four factors or more) are needed, estimation will be completed via the Metropolis Hastings – Robbins Monro algorithm (e.g., Cai, 2010) and standard errors will be estimated via the Louis formula (Louis, 1982) with 2500 iterations.

The overall fit of models examined in flexMIRT will be evaluated using the Tucker-Lewis Index (TLI, Tucker & Lewis, 1973) and the limited-information M_2-based root mean square error of approximation (RMSEA, Maydeu-Olivares et al., 2011). The cut-off for adequate fit of the TLI will be 0.97 (Cai et al., 2023) with values at or greater than 0.97 indicating good fit. For the RMSEA, the appropriate cutoff can vary based on a variety of factors, including number of items and number of categories presented for item responses. Given that CIFA analyses often involve collapsing categories to account for sparseness in responses, the varying numbers of potential item response categories make firm guidelines for the RMSEA value unclear. As a result, a cut-off value of less than 0.05 (Monroe & Cai, 2015) will be used, though failing to meet this criterion will not be considered exclusively disqualifying for the CIFA model.

For the MiCOAS 7d and 14d, 4 unidimensional models (1 per *a priori* measure) will be examined. As noted previously, the 4 measures are intended to assess PF (Figure 2; full item text in Table 4), SCF (Figure 3; full item text in Table 5), SF (Figure 4; full item text provided in Table 6), and EF (Figure 5 ; full item text provided in Table 7).

Figure 2. Path diagram for the *a priori* unidimensional model for the MiCOAS 7/14-day PF measure

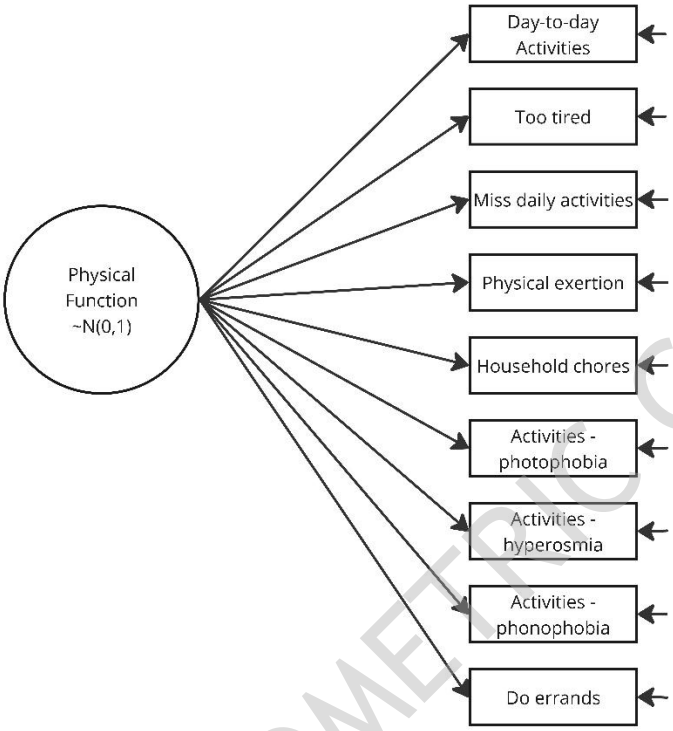


Figure 3. Path diagram for the *a priori* unidimensional model for the MiCOAS 7/14-day SCF measure

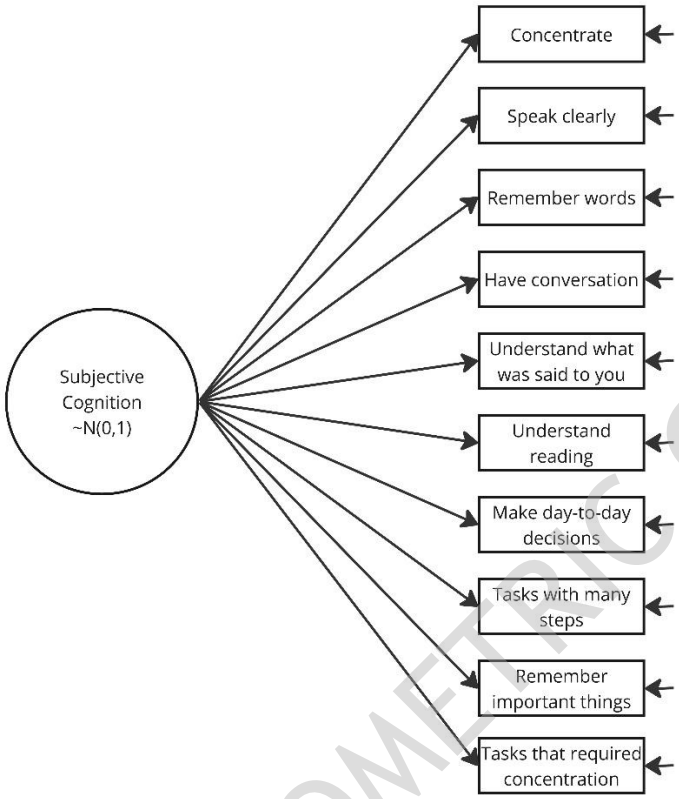


Figure 4. Path diagram for the *a priori* unidimensional model for the MiCOAS 7/14-day SF measure

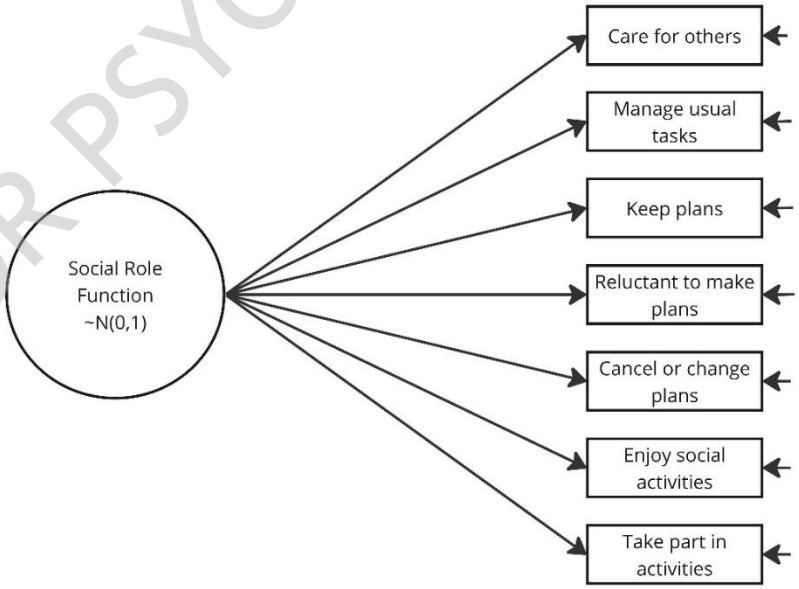
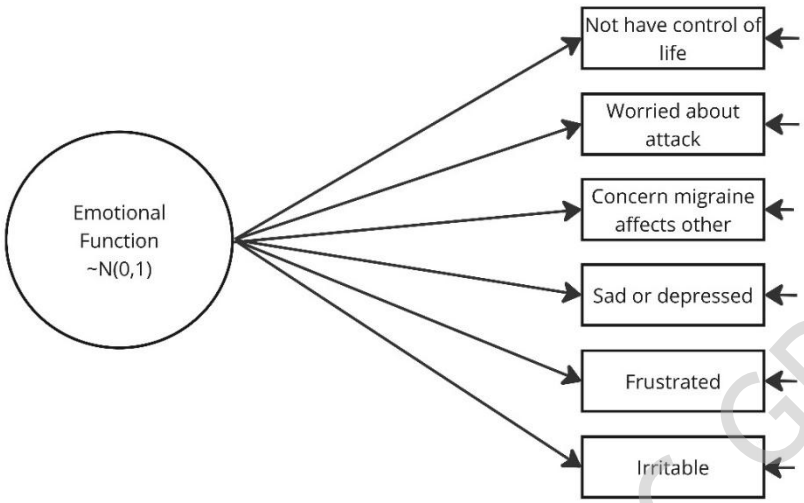


Figure 5. Path diagram for the *a priori* unidimensional model for the MiCOAS 7/14-day EF measure



For each examined CIFA model, factor loadings and overall model fit values will be reported (e.g., Table 10).

Table 10. CFA factor loadings and overall model fit values of the MiCOAS 7d PF measure items at Baseline

Item content	Loading
Was it difficult for you to do your usual day-to-day activities?	0.xx
How often were you too tired to do your regular daily activities?	0.xx
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	0.xx
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	0.xx
Was it difficult for you to do your regular household chores?	0.xx
Was it difficult for you to do activities because you were sensitive to light?	0.xx
Was it difficult for you to do activities because you were sensitive to smells?	0.xx
Was it difficult for you to do activities because you were sensitive to sound?	0.xx
Was it difficult for you to do errands?	0.xx
Model fit	
RMSEA	0.xx
TLI	0.xx

Similar tables provided for the other *a priori* MiCOAS 7d measures and all MiCOAS 14d measures

If the confirmatory models (or minorly adjusted versions of the *a priori* models, which is defined as removal of items or an alternate model that still produces one interpretable summary score, such as a bifactor model to account for method factors) do not provide acceptable fit for any of the recall-based measures, EIFA models will be explored, separately, for the items of each *a priori* domain that failed to achieve acceptable fit. The determination regarding the number of factors to extract will be based on results from parallel analysis (e.g., Liu & Rijmen, 2008), supplemented with an examination of the eigenvalues and the scree plot constructed from those eigenvalues from each examined correlation matrix. This examination will result in a target EIFA model (for example, a 4-factor model) and a model with one and two fewer dimensions (e.g., a 2-factor and 3-factor model) will also be reported. For EIFA models with more than one factor, oblique Crawford Ferguson-quartimax rotation (e.g., Browne, 2001) will be used to rotate initial factor extraction values to more readily interpretable solutions in flexMIRT. Factor loadings for each item and factor intercorrelations will be reported for each examined EIFA model (Table 11).

Table 11. Factor loadings and factor intercorrelations for an m-dimensional EFA model

Item content		F1	…	F_last
Was it difficult for you to do your usual day-to-day activities?		0.xx	0.xx	0.xx
How often were you too tired to do your regular daily activities?		0.xx	0.xx	0.xx
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.		0.xx	0.xx	0.xx
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?		0.xx	0.xx	0.xx
Was it difficult for you to do your regular household chores?		0.xx	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to light?		0.xx	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to smells?		0.xx	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to sound?		0.xx	0.xx	0.xx
Was it difficult for you to do errands?		0.xx	0.xx	0.xx
Factor correlations				
	F1	1		
	…	0.xx	1	
	F_last	0.xx	0.xx	1

The decision to select a preferred EIFA model (i.e., number of factors/factor pattern) will be based on statistical considerations (e.g., item factor loadings values) as well as substantive considerations (e.g., interpretability of the resulting factors). At this point, items with low factor loadings or other undesirable psychometric properties (e.g., ceiling or floor effects) will be candidates for removal from the measure, in consultation with clinical experts in migraine. The preferred EIFA model will then be tested within the CIFA framework.

Item Response Theory Analyses

If a well-fitting and substantively meaningful CIFA model is found for any of the MiCOAS item sets (3 measures from each recall-based item set) this model will be reparameterized within the IRT framework to examine the psychometric performance of the individual items and the measures' scores in more depth. IRT parameters will be obtained, using the categorical data analysis methods and implementations described previously. Item parameters and overall model fit and item fit statistic values (Table 12), item trace lines curves (Figure 6), and test reliability functions (Figure 7) will be reported for each measure (as available, given the dimensionality of the final model).

Table 12. IRT item parameters and item-level fit index values for MiCOAS 7d PF measure

Item content	a/slope	b1	b2	b3	b4	S - X^2 value (df), p
Was it difficult for you to do your usual day-to-day activities?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
How often were you too tired to do your regular daily activities?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do your regular household chores?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do activities because you were sensitive to light?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do activities because you were sensitive to smells?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do activities because you were sensitive to sound?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Was it difficult for you to do errands?	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx

Figure 6. Example item trace line plot for a 5-category GRM item

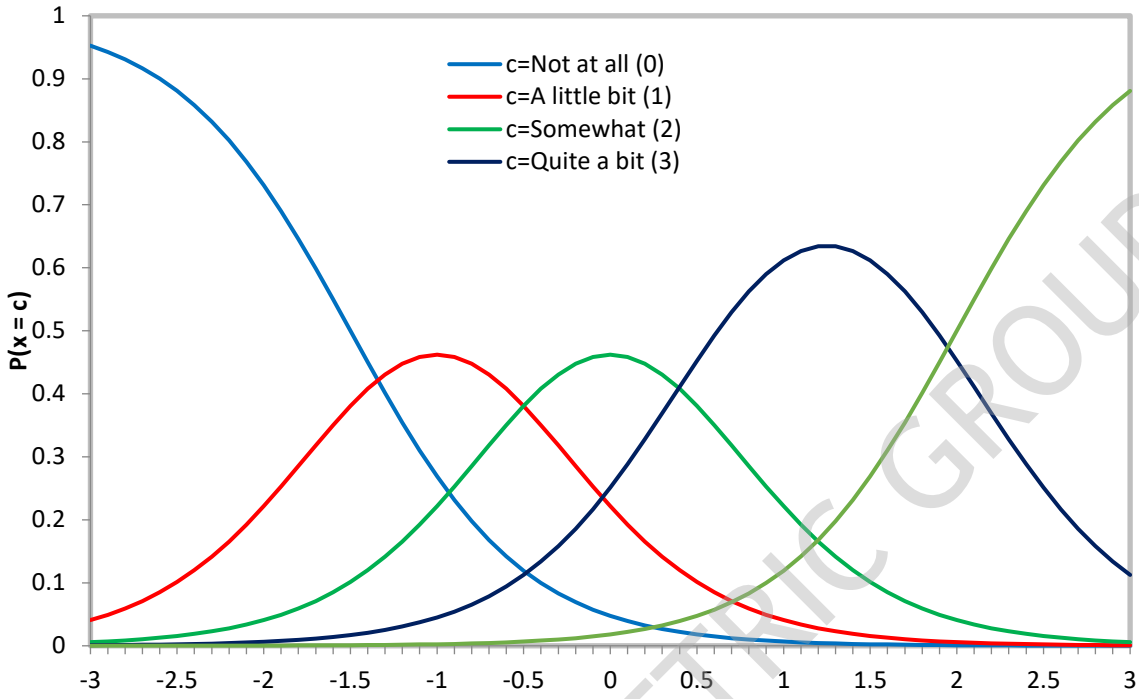
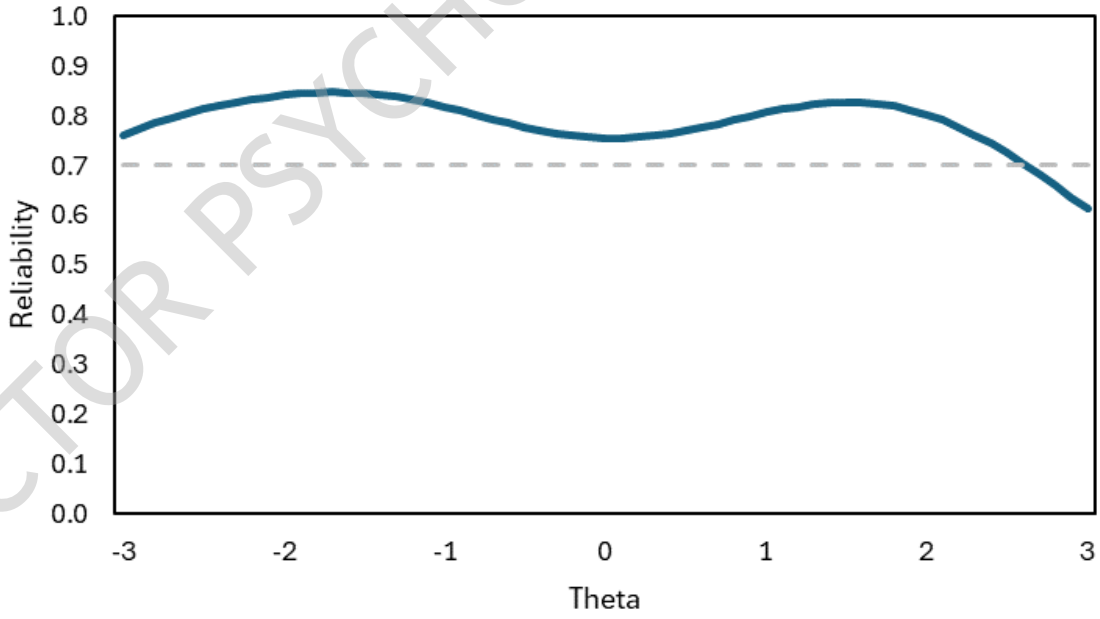


Figure 7. Example test reliability function plot



Additionally, for each recall-based measure, differential item functioning (DIF) analyses will be completed for groups defined by select demographic and descriptor variables. Currently, this includes chronic versus episodic migraine (CM and EM, respectively), and headache day versus non-headache day (for DA only). DIF will be implemented using the two-stage process described in Woods et al. (2013), in which a "sweep" analysis is first run to identify strongly invariant items across groups and then a more typical anchored analysis is used for the final DIF analysis, using the invariant items defined in the "sweep" step as anchor items. DIF will be tested using the Langer-improved Wald test (Langer, 2008). Overall Langer-improved Wald test p-values will be judged against Benjamini-Hochberg (1995) corrected p-values to prevent Type I errors due to multiple testing. If the overall DIF test for an item is found to be statistically significantly different across groups, examinations into the meaningfulness/severity of the detected DIF (e.g., Hansen et al., 2014) using the weighted area between the curve (wABC) to quantify differences between groups will be completed prior to final determinations regarding flagging an item for non-trivial DIF. Items with a wABC value greater than 0.30 (Hansen et al., 2014) will be considered to exhibit non-ignorable DIF.

Reliability

Reliability is a measure of the extent of consistency or repeatability of scores obtained from a given model. The recall-based MiCOAS measures will be assessed for both internal consistency and test-retest reliability.

Internal Consistency Reliability

Internal consistency reliability evaluates individual items in comparison with one another for their ability to give consistently appropriate results. To investigate internal consistency of the items contributing to the MiCOAS recall-based measures separately (that is, 4 recall-based measures from both the MiCOAS 7d and MiCOAS 14d item sets), CTT analyses, (i.e., coefficient alpha, alpha with item *i* removed, and item-total correlations) will be reported using data from the previously described Baseline sample and values will be reported similar to Table 13.

Table 13. Coefficient alpha, item-total correlations, alpha-if-deleted for MiCOAS 7d PF measure at Baseline

Parameter		Value
N		xxx
Overall alpha		0.xx
Item	Item-total correlation	Alpha-if-deleted
Was it difficult for you to do your usual day-to-day activities?		
How often were you too tired to do your regular daily activities?	0.xx	0.xx
How often did you have to miss your regular daily activities because of migraine? Include times that you stopped early or started late.	0.xx	0.xx
Was it difficult for you to do usual activities that require physical exertion, like going up stairs?	0.xx	0.xx
Was it difficult for you to do your regular household chores?	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to light?	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to smells?	0.xx	0.xx
Was it difficult for you to do activities because you were sensitive to sound?	0.xx	0.xx

Similar tables provided for other MiCOAS 7d measures and all MiCOAS 14d measures

If a multidimensional solution is found as the preferred model for any of the MiCOAS recall-based measures, the items contributing to each dimension will be submitted to CTT analysis separately. Additionally, the marginal (distributional) reliability value from the IRT-based analyses will be reported. A value of 0.70 will be used as the minimum overall coefficient alpha and the minimum marginal reliability value to demonstrate satisfactory internal consistency (e.g., Nunnally, 1978).

Test-retest reliability

Test-retest reliability will be assessed for the each of the MiCOAS recall-based measure scores (Table 14). For the MiCOAS 7d, Baseline and Week 1 assessments from Cohort 3 will be used. For the MiCOAS 14d, Baseline and Week 2 assessments from Cohorts 1 and 2 will be used. Test-retest will be evaluated using uncorrected Pearson correlations as well as intraclass correlation coefficients (ICCs) from a two-way mixed effect model with absolute agreement for single measures (McGraw & Wong, 1996). ICCs will be judged against thresholds established for interpreting intraclass correlations (i.e., values less than 0.40 indicate “poor,” 0.40 to 0.75 as “fair to good;” 0.75 to 0.90 as “good,” [Fleiss, 1986]).

Because the variability and change often seen in observational migraine studies as well as in placebo groups of randomized clinical trials, test-retest analyses will be performed on a “stable” subset of participants. For the MiCOAS 7d item sets, stability is defined as no change in the PGIS response from Baseline (Fay 0) to Week 1 (Day 7) using data from Cohort 3. For the MiCOAS 14d item sets, stability is defined as no change in PGIS responses between Baseline (Day 0) and Week 2 (Day 14) using data from Cohort 1 and 2. If an insufficient number of stable participants (fewer than 30) are identified by these definitions, results will be reported with a caution regarding overinterpretation of values derived from small samples.

Table 14. Test-retest values for MiCOAS recall-based measure scores using a stable subset of participants.

Measure (Time period)	Pearson correlation	ICC
MiCOAS 7d (Baseline [Day 0] to Week 1 [Day 7])		
PF	0.xx	0.xx
SCF	0.xx	0.xx
SF	0.xx	0.xx
EF	0.xx	0.xx
MiCOAS 14d (Baseline [Day 0] to Week 2 [Day 14])		
PF	0.xx	0.xx
SCF	0.xx	0.xx
SF	0.xx	0.xx
EF	0.xx	0.xx

VALIDATION ANALYSES

Several relationships with other assessments will be examined using the MiCOAS recall-based measure scores. Given the exploratory nature of these analyses, no family-wise error rate corrections will be made to account for multiple testing. Additionally, the number and interpretation of scores from the MiCOAS recall-based item sets/measures will depend on the results of the factor analyses; the expectations stated below are based on the assumption that the *a priori* unidimensional models of 4 measures per questionnaire are supported and will produce individual scores (e.g., PF, SCF, SF, EF measure scores from both MiCOAS 7d or MICOAS 14d).

The first validity evidence to be presented will be descriptive statistics of each MiCOAS measure's scores over the target assessments. Available sample size, mean, median, minimum, maximum, and SD for each score will be reported for each recall-based measure at each target timepoint. As stated previously, for the MiCOAS 7d measures administered in Cohort 3, target timepoints are Baseline (Day 0), Week 1 (Day 7), Week 2 (Day 14), Week 3 (Day 21), and Week 4 (Day 28); Table 15). For the MiCOAS 14d versions, target timepoints are Baseline (Day 0), Week 2 (Day 14), Week 4 (Day 28), Week 6 (Day 42), and Week 8 (Day 56) and will use data from both Cohort 1 and 2.

Table 15. Descriptive statistics of the MiCOAS 7d measures by target timepoint

Timepoint/ Descriptive	PF	SCF	SF	EF
Baseline (Day 0)				
N	xx	xx	xx	xx
M(SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx
Week 1 (Day 7)				
N	xx	xx	xx	xx
M(SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx
Week 2 (Day 14)				
N	xx	xx	xx	xx
M(SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx
Week 3 (Day 21)				
N	xx	xx	xx	xx
M(SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx
Week 4 (Day 28)				
N	xx	xx	xx	xx
M(SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx	xx, xx

Similar table provided for MiCOAS 14d measure scores in Cohort 1 and 2 at target timepoints

Convergent/Discriminant Correlation Evidence

Convergent and discriminant evidence provides information regarding the extent to which the scores of interest exhibit relationships as expected with scores from other theoretically distinct or similar scales or variables. Convergent evidence asserts that two measures representing theoretically related constructs should be meaningfully correlated. In contrast, discriminant evidence should demonstrate that measures representing theoretically unrelated or more distal constructs should produce near-zero or small observed correlations with the scores of interest. In addition to the point estimates of evaluated correlations, the overall pattern of results requires interpretation.

Correlational methods used to examine the following relationships will vary depending on the data characteristics of the variables being correlated (e.g., if MiCOAS measure scores are correlated with a reference measure with limited response categories [7 or fewer], Spearman correlations will be reported; if the target scores are correlated with a continuous or large-category ordinal reference measure, Pearson correlations will be reported). The expected direction of correlations between the 7- and 14-day recall versions of the MiCOAS PRO measure scores and reference variables are specified in Table 16 and observed values will be reported within each target timepoint (e.g., Table 17).

Table 16. Expected direction of correlations among MiCOAS 7d and MiCOAS 14d scores and reference variables.

Variable	PF	SF	EF	CF
Age	-0.2 to 0.2	0 to 0.3	0 to 0.3	0 to 0.3
Height	-0.2 to 0.2	-0.2 to 0.2	-0.2 to 0.2	-0.2 to 0.2
MOS-Cog-R	-0.3 to 0	-0.3 to 0	-0.3 to 0	-0.5 or more extreme
HIT-6 Total score	0.4 or greater	0 to 0.3	0 to 0.3	0.1 to 0.4
Item: Limit ability to concentrate	0 to 0.3	0 to 0.3	0 to 0.3	0.4 or greater
MFIQ v 2.0				
Physical function	0.3 to 0.8	0.1 to 0.4	0.1 to 0.4	0 to 0.3
Usual activities	0.4 or greater	0.1 to 0.4	0.1 to 0.4	0 to 0.3
Social function	0.1 to 0.4	0.4 or greater	0.4 or greater	0 to 0.3
Emotional function	0 to 0.3	0.3 to 0.8	0.3 to 0.8	0 to 0.3
Item: Difficult...activities required concentration	0 to 0.3	0 to 0.3	0 to 0.3	0.4 or greater
MSQ v2.1				
Role function - restrictive	-0.5 to -0.3	-0.5 to -0.3	-0.5 to -0.3	-0.3 to 0
Role function - preventive	-0.5 to -0.3	-0.5 to -0.3	-0.5 to -0.3	-0.3 to 0
Emotional function	-0.3 to 0	-0.5 to -0.3	-0.5 or more extreme	-0.3 to 0
Item: Limit ability to concentrate	-0.3 to 0	-0.3 to 0	-0.3 to 0	-0.5 to -0.3
MMD	0.3 to 0.5	0.2 to 0.5		0.2 to 0.5

Note. PF = Physical function. SF = Social function. EF = Emotional function. CF = Cognitive function.

Table 17. Correlations of MiCOAS 7d scores with reference measures at Baseline (Day 0)

	PF		SCF		SF		EF	
Reference measure	N	r	N	r	N	r	N	r
MOS-Cog-R 7d	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
HIT-6 (4-week recall)	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Too tired to do work	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MSQ v2.1 (4-week recall)								
Role Function - Restrictive	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Role Function- Preventative	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Difficult...activities required concentration	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MFIQ v2 (7-day recall)								
Physical Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Usual Activities	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Social Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***

Similar tables provided for other MiCOAS 7d target timepoints and all MiCOAS 14d scores at each target timepoint

Known-Groups Evidence

Known-groups analyses are examined to determine the extent to which scores of interest reflect expected differences between clinically meaningful and conceptually distinct groups.

For the recall-based MiCOAS measures known-groups analyses, clinically meaningful groups will be formed from the following reference variables:

- PGIS response categories: it is expected that as PGIS categories become more severe (from None to Very Severe) the recall-based MiCOAS scores will also increase
- PGIC response categories: it is expected that as PGIC response categories increase (from Very Much Worse to Very Much Improved) the recall-based MiCOAS measures scores will also increase
- HIT-6 groups: it is expected that as HIT-6 severity category increases (from Little to No Impact up through Severe impact), the recall-based MiCOAS scores will also increase.

For the scores from the MiCOAS 7d and 14d measures, cross-sectional known-groups analyses within each target timepoint are planned (e.g., Table 18). For these continuous scores, group sample size, means, and SDs will be reported and analysis of variance (ANOVA) will be used to examine differences in each measure's scores across clinically meaningful groups. For all known-groups analyses, no post-hoc tests to probe statistically significant overall ANOVA effects will be conducted and, to prevent the over-interpretation of results derived from small n values, analysis results will only be reported if groups constructed from the reference variables have at least 25 observations.

Table 18. Known-groups analyses of MiCOAS 7d PF measure scores by target timepoint

	Baseline (Day 0)			Week 1 (Day 7)			Week 2 (Day 14)			Week 3 (Day 21)			Week 4 (Day 28)		
Anchor/Groups	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
PGIS															
None	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Mile	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Moderate	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Severe	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Very severe	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Overall Effect (ANOVA)	F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx		
PGIC															
Very much improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Much improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Minimally improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
No change				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Minimally worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Much worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Very much worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Overall Effect (ANOVA)				F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx		
HIT-6 groups	xxx	x.x	x.x												
Little to no impact	xxx	x.x	x.x										xxx	x.x	x.x
Moderate impact	xxx	x.x	x.x										xxx	x.x	x.x
Substantial impact	xxx	x.x	x.x										xxx	x.x	x.x
Severe impact	xxx	x.x	x.x										xxx	x.x	x.x
Overall Effect (ANOVA)	F(x,xxx)=xx.x, p=0.xx												F(x,xxx)=xx.x, p=0.xx		

Similar table provided for other MiCOAS 7d scores and all MiCOAS 14d scores at target timepoints

Ability to Detect Change

The ability to detect change over time is critical for understanding the adequacy of measures for clinical research. In short, if patients have changed over the course of the study, that change will be reflected in the values of scores that are valid for their intended purpose. Given the observational nature of the quantitative study, it is unknown what level of change over time will be realized in the data, but in consideration of the fact that migraine observational studies and placebo groups in migraine trials do tend to change, the ability to detect change of each of the MiCOAS measures' scores will be investigated.

The assessment of ability to detect change will be examined by creating change scores for the recalled-based MiCOAS measures (7d and 14d) scores defined as change = Baseline – Week 4 (Day 28). Correlations among the change in reference variables and the change in each MiCOAS measure's score will be examined. The direction and magnitude of the specified correlations are expected to mirror the relationships described in the convergent and discriminant evidence section above and will be reported in a format similar to Table 19.

Table 19. Correlations of change in MiCOAS 7d scores with change in reference variables (Baseline to Week 4 [Day 28])

	PF		SCF		SF		EF	
Reference measure	N	r	N	r	N	r	N	r
MOS-Cog-R 7d	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
HIT-6 (4-week recall)	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Too tired to do work	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MSQ v2.1 (4-week recall)								
Role Function - Restrictive	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Role Function- Preventative	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Difficult...activities required concentration	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MFIQ v2 (7-day recall)								
Physical Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Usual Activities	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Social Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***

Similar tables provided for MiCOAS 14d versions

MEANINGFUL SCORE REGION ANALYSES

Meaningful score regions (MSRs) will be estimated for MiCOAS summary score using the baseline data for the recall-based measures. This analysis will employ an anchor-based approach using responses on the PGIS (a priori primary anchor) to identify MSR for the each of the MiCOAS recall-based scores. Using the responses of the PGIS item at Baseline, summary values that describe the distribution of the MiCOAS scores within each

response group will be constructed (Table 20), along with boxplots to visually depict these distributions (top section of Figure 8).

Table 20. Summary of Baseline MiCOAS 7d scores by PGIS item responses at Baseline

PGIS at Baseline (Day 0)	n	Min	Q1	Mean	Median	Q3	Max	SD
			PF					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			SCF					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			SF					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			EF					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

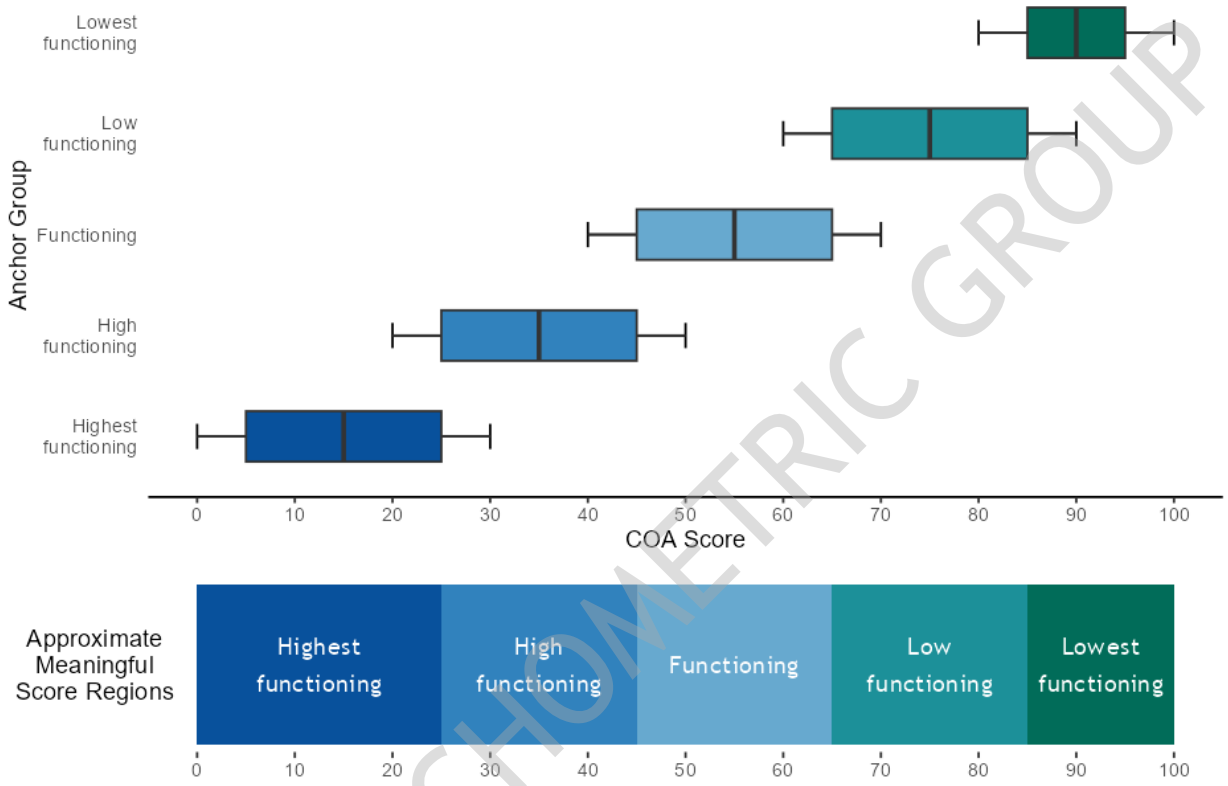
Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity.

Similar table provided MiCOAS 14d scores

To define the cut-values between response categories (based on FDA suggestion for such requests in other projects), the 75th percentile (Q3) of each MiCOAS summary score of the lower response category is used to define the cut-value between two PGIS response categories. For example, if the “None” PGIS response is found to have a range of a given MICOAS score (which is planned to be on a t-distribution with a mean of 50 and a SD of 10) from 25 to 41 with a score of 34 found to be in the 75th percentile of the distribution of scores, the value of 34 is used to define the threshold to separate the “None” from “Mild” score ranges. A similar process will be used to define the cut-values between subsequent PGIS response categories, resulting in non-overlapping categories of each MiCOAS score for groups defined by PGIS ratings of severity, which can

eventually be used to assist in the interpretation of observed differences in treatment group scores (bottom section of Figure 8).

Figure 8. Example box plots of target COA scores by primary anchor variable categories at Baseline.



MEANINGFUL WITHIN-PERSON CHANGE/MEANINGFUL SCORE DIFFERENCE (MSD) THRESHOLD ESTIMATES

As noted earlier, it is unknown what level of change over time will be realized in the data, but in consideration of the fact that migraine observational studies and placebo groups in migraine trials do tend to change and the importance of thresholds to aid in the interpretation of scores, exploratory analyses to begin to investigate meaningful within-person change threshold value for each of the recall-based MiCOAS measures' scores will be investigated. A within-person meaningful score difference (MSD_w) threshold sets the minimum amount of individual change necessary for a patient to experience a "meaningful" change in their condition over time; such thresholds are important tools for understanding and interpreting the observed change in an outcome of interest. For the MiCOAS 7d and 14d scores, improvement and worsening MSD_w thresholds will be estimated (if change over time in a sufficient number of participants is observed). Both anchor- and distribution-based methods will be used to investigate change in the developing recall-based MiCOAS scores and obtain candidate MSD_W threshold values for each of the recall-based MiCOAS scores (if possible).

Anchor-based methods

To establish the suitability of potential anchors, correlations among change scores (defined for each measure as previously noted in the Ability to Detect Change section) and the candidate anchor variables (change in the PGIS, PGIC responses at Week 4/Month 1, change in the HIT-6 total score) will be examined against the change in each of the MiCOAS measures' scores from each of the three questionnaires. *A priori*, the change in the PGIS item is considered the primary anchor for the recall-based MiCOAS measures and meaningful change is defined as an improvement or worsening of 1 response category between Baseline and Week 4 (Day 28) (e.g., Severe to Moderate for improvement, Moderate to Severe for worsening). For the PGIC, a response of "Minimally better" is considered meaningful improvement and a response of a "Minimally worse" is considered meaningful worsening. For the change in the HIT-6 from Baseline (Day 0) to Week 4 (Day 28), a decrease of 6 is considered meaningful improvement (Houts et al., 2020); there is no empirically derived worsening threshold for the HIT-6 total score, but an increase of 6 (consistent with the improvement threshold) will be considered meaningful worsening.

For candidate anchors to be suitable for use, it is generally desirable that the target change scores be correlated with the potential anchor at least 0.3 (e.g., Coon & Cook, 2018; Revicki et al., 2008). Spearman correlations between the change in the MiCOAS recall-based scores and change in the potential anchors will be reported to allow for a determination of suitability of use as an anchor with each MiCOAS score.

To examine the extent to which the observed change over time is influenced by baseline status, a table of key percentiles of the recall-based MiCOAS change scores, stratified by participants baseline status on the PGIS will be constructed (Table 21), similar to FDA's patient-focused drug development guidance 4's Table 1 (pg. 23) but including all participants.

Table 21. Key percentiles of MiCOAS 7d PF change scores by baseline PGIS response for all participants

		Change in MiCOAS score, by key percentile				
PGIS at Baseline	n (%)	10th	25th	50th	75th	90th
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity.

Similar tables constructed for other MiCOAS 7d scores and all MiCOAS 14d scores

Additionally, to understand the change in those participants who meaningfully improved or worsened, a similar table of key percentiles for only those participants who meaningfully changed (Table 22) will be reported.

Table 22. Key percentiles of MiCOAS 7d PF change scores by Baseline (Day 0) PGIS response for participants who improved or worsened 1-category on the PGIS between Baseline (Day 0) and Week 4 (Day 28)

		Change in MiCOAS score, by key percentile				
PGIS at Baseline (Day 0)	n (%)	10th	25th	50th	75th	90th
			Meaningfully Improved			
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			Meaningfully Worsened			
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity. PGIS responses of None in the meaningfully improved section and Very severe in the meaningfully worsened sections have been excluded as it is not possible to improve/worsen from these responses.

Similar tables constructed for all other MiCOAS 7d scores and all MiCOAS 14d scores

For the anchors that are acceptable for use with each of the MiCOAS recall-based scores, all categories will be used to define groups for determining candidate MSD_w values for improvement. Additionally, to ensure that illogical threshold values are not chosen (e.g., for the PGIC item, the "No change" group having a larger MiCOAS scores change than the "Minimally improved" group), descriptive statistics of the MiCOAS recall-based change scores at each level of each anchor item will be reported (e.g., Table 23).

Table 23. Descriptive statistics of MiCOAS 7d change scores by anchor variables

	PF					SCF					SF					EF				
Anchor variable / Levels of anchor variable	n	M	Median	SD	Min, Max	n	M	Median	SD	Min, Max	n	M	Median	SD	Min, Max	n	M	Median	SD	Min, Max
Change in PGIS response (Baseline - Week 4 [Day 28])																				
3+ category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
2 category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
1 category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
1 category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
2 category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
3+ category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
PGIC response (Week 4 [Day 28])																				
Very much improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Much improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Minimally improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Minimally worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Much worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Very much worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Change in HIT-6 (Baseline - Week 4 [Day 28])																				
Improvement responder (+6 or more better)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change (-6 to 6)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Worsening responder (-6 or more worse)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx

Note. n = group size. M = Mean. SD = Standrd deviation. Min = minimum. Max = Maximum.

Similar table provided for MiCOAS 14d scores

The approach for identifying thresholds for interpretation will be the presentation of empirical cumulative distribution functions (eCDFs) and empirical probability distribution functions (ePDFs). eCDFs plot the cumulative proportion of subjects reporting change scores on a COA along the range of the COA change scores on the x-axis. eCDFs will present separate curves for each level of change on the anchor (e.g., Figure 9). While there is no universal guideline for interpreting such eCDFs, the location at which the meaningful deterioration curve reaches 50% on the y-axis is a possible location for an interpretable threshold. The ePDF curves (e.g., Figure 10) are useful in aiding the interpretation of eCDF curves. While presenting the same information as the eCDFs, ePDF curves may provide a more interpretable overview of the shape, dispersion, and skewness of the distribution of the change from baseline in the endpoint of interest across various anchor categories.

Figure 9. Example eCDF for determining meaningful within-person change thresholds

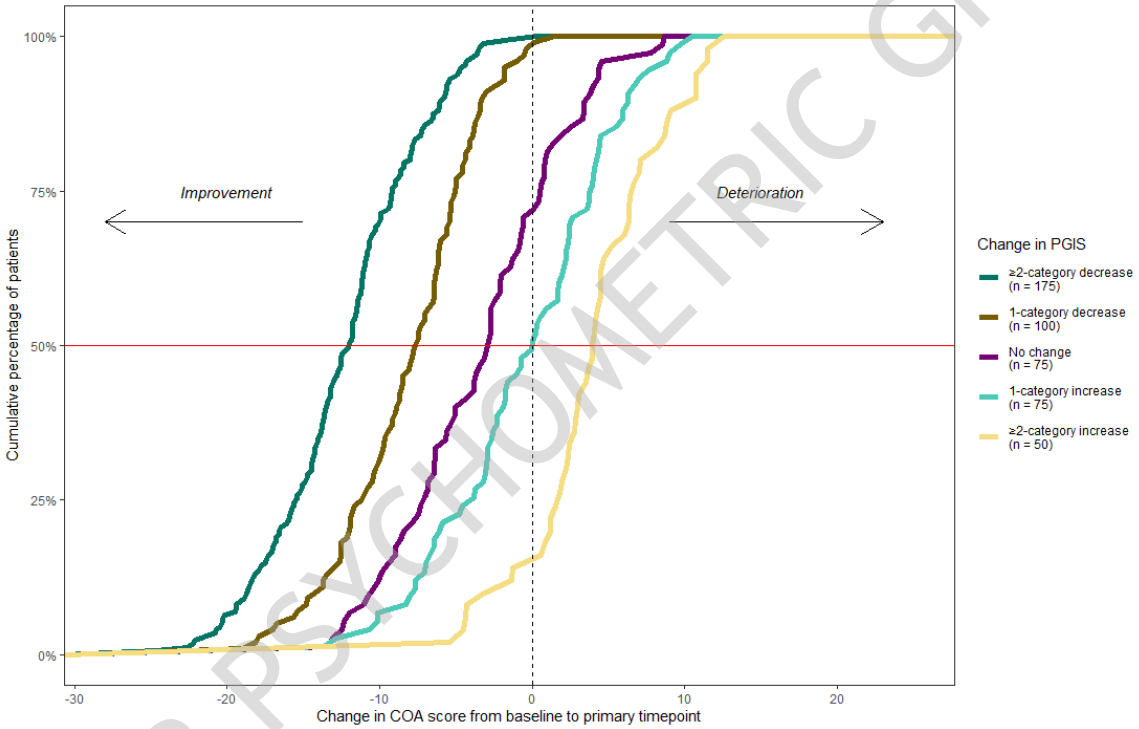
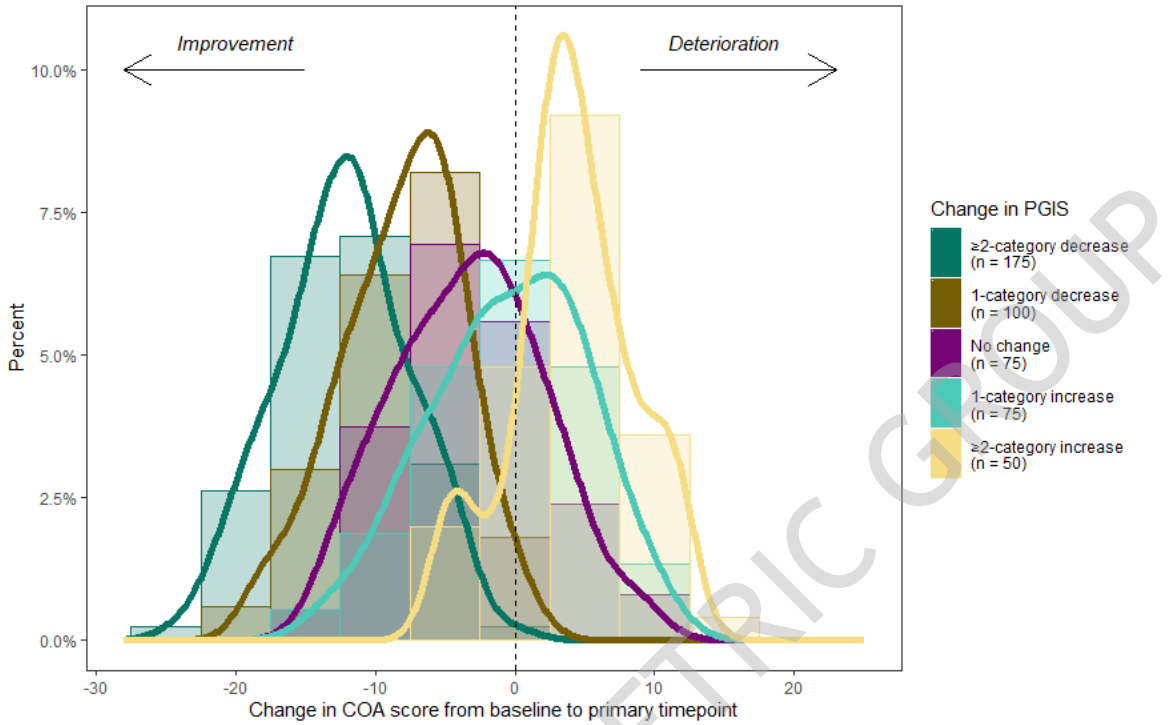


Figure 10. Example ePDF for determining meaningful within-person change thresholds



Distribution-based methods

For the distribution-based estimates, 2 typical values will be reported for each of the MiCOAS PRO measure scores: 1) ½ SD of Baseline scores and 2) the standard error of measurement (SEM) at Baseline. The previously found Baseline internal consistency alpha value will be used as the reliability estimate in calculating the SEM, using the formula $\text{SEM} = \text{SD} * \sqrt{(1 - \text{reliability})}$. While these distribution-based values are not patient-centered, they will be used to set a minimum value that any anchor-based suggested MSD_w threshold value must exceed to ensure that a change over time that could solely be attributable to measurement error/noise in the MiCOAS PRO measures scores is not considered "meaningful".

Triangulation

The candidate MSD_w values based on the anchor- and distribution-based methods described above will be collected into a table for improvement (e.g., Table 24) and, separately for worsening. Tabled values will be considered to identify a value or a range of values that may be used for interpreting change scores (i.e., triangulation). Distribution-based values, as noted above, will be used to set a minimum change value, while results from the anchor-based methods will be evaluated and weighted with consideration of the content relevance and strength of correlation of the anchor with the MiCOAS recall-based scores during the triangulation process to establish a final MSD_W value/range of values.

Table 24. Candidate anchor- and distribution-based improvement MSD_w values for MiCOAS recall-based scores

	Anchor-Based			Distribution-based	
	Change in PGIS	PGIC	Change in HIT-6 Total		
Version/measure	+1 Category Median	Improved Median	Improved Median	1/2 SD	SEM
MiCOAS 7d recall					
PF	x.xx	x.xx	x.xx	x.xx	x.xx
SCF	x.xx	x.xx	x.xx	x.xx	x.xx
SF	x.xx	x.xx	x.xx	x.xx	x.xx
EF	x.xx	x.xx	x.xx	x.xx	x.xx
MiCOAS 14d recall					
PF	x.xx	x.xx	x.xx	x.xx	x.xx
SCF	x.xx	x.xx	x.xx	x.xx	x.xx
SF	x.xx	x.xx	x.xx	x.xx	x.xx
EF	x.xx	x.xx	x.xx	x.xx	x.xx

Note. Similar table provided for worsening (provided sufficient data).

METHODS FOR DA MICOAS MEASURES

For clarity, the item content and response scale of each of the *a priori* DA measures are reported below in Table 25 to Table 27.

Table 25. Items included in the MiCOAS DA Symptom measure

Item content	Response scale
Did you have a headache or head pain?	Severity
Aura was	Severity
Dizziness or vertigo was	Severity
Nausea was	Severity
Neck pain was	Severity
Sensitivity to light was	Severity
Sensitivity to sound was	Severity
Vomiting was	Severity
Which one of the following symptoms other than pain was most bothersome to you? (Select one)	

Table 26. Items included in the MiCOAS DA PF measure

Item content	Response scale
Tiredness was	Severity
Bending forward was	Difficulty
Getting read for the day (e.g., get showered and dressed) was	Difficulty
Keeping your plans for the day	Difficulty
Moving your body was	Difficulty
Moving your head was	Difficulty
Strenuous physical activity was	Difficulty
Walking was	Difficulty

Table 27. Items included in the MiCOAS DA SMA measure

Item content	Response scale
Concentrating was	Difficulty
Mental slowness or fogginess was	Severity
Remembering the correct words for things was	Difficulty
Speaking clearly was	Difficulty
Thinking clearly was	Difficulty

ITEM-LEVEL DESCRIPTIVE STATISTICS

Descriptive statistics and an observed frequency table for each of the MiCOAS DA will be reported for the first 7 days after Baseline (Day 0), the 7 days before Week 4 (including Day 28), the 7 days before Week 8 (including Day 56), and for a random selection of one day per participant from the first 28 days (Day 1 to Day 28). Tables will be formatted similarly to Table 28.

PSYCHOMETRIC ANALYSES

Analyses to investigate the psychometric properties of the items contributing to the MiCOAS DA will be completed.

In consideration of the intensive longitudinal nature of the MiCOAS DA (that is, daily assessments for up to 2 months) and the serial correlation that can occur between temporally close assessments (e.g., today and tomorrow versus today and 15 days later), from the assessments in Month 1 (Day 1 through Day 28) one day per participant (in which at least 50% of the DA items have valid responses) will be randomly selected from the available assessments. These data will be used for the polychoric correlations, dimensionality assessment (EIFA and CIFA) and IRT analyses to investigate the psychometric properties of the MiCOAS DA items.

Table 28. Descriptive summary and frequency table of MiCOAS DA items, Day 1

Content	N	M	Mode	SD	Cat0	Cat1	Cat2	Cat3	Cat4	Cat5	Cat6	Cat7	Cat8
Did you have a good day	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---	---
Did you have a migraine day	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---	---
Did you have a headache or head pain	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
How long did the headache last	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Pain worse with activity	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---
Throbbing/pounding head pain	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---
Pain only or worse on one side	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---
Take medication to alleviate pain	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---	---
Did you experience any of the following													
Nausea	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Vomiting	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Sensitivity to light	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Sensitivity to sound	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Neck pain	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Aura	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Mentally slow or foggy	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Tired	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---	---
Which symptom other than pain was most bothersome?	xxx				n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)
Were any of the following difficult?													
Ability to think clearly	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to concentrate	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Speak clearly	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)				
Remember words	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)				
Ability to get ready for the day	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to keep your plans	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to move your body	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to walk	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to move your head	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to bend forward	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---
Ability to do strenous physical activity	xxx	xx.x	xx.x	xx.x	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	n (xx.x%)	---	---	---	---

For symptom items, 0 = Did not experieence, 1 = Mild, 2 = Moderate, 3 = Severe. For difficulty items, 0 = did not experience, 1 = a little difficult, 2 = Somewhat difficult, 3 = Very difficult, and 4 = Unable to do. For Most bothersome item, 1 = Aura, 2 = Dizziness or vertigo, 3 - mentral slowness or fogginess, 4 = Nausea, 5 = Neck pain, 6 = Sensitivity to light, 7 = Sensitivity to sound, 8 = Tired, 9 = Vomiting, 10 = Other

Similar tables provided for all other target days and for the randomly sampled days of the MiCOAS DA.

As an initial tool to investigate how the items of each of the various MiCOAS measures are related to one another within each questionnaire, polychoric correlations among the (collapsed, as needed) items of the MiCOAS DA will be reported (Table 29). For the item that is included in the DA but not assigned to any *a priori* measure, polychoric correlation values will be reviewed to determine which, if any, measure this item could be assigned to.

Factor Analytic Modeling

For the MiCOAS DA items, which have limited response categories making continuous data analysis methods unwise (e.g., Wirth & Edwards, 2007), item factor analytic modeling (confirmatory and exploratory, if necessary) will be conducted in flexMIRT 3.7 (Cai, 2024). In flexMIRT, a full-information categorical estimation method, specifically Bock-Aitkin expectation-maximization (BAEM) estimation (Bock & Aitkin, 1981), will be employed. Standard errors (SEs) of item parameters will be estimated using the supplemented EM algorithm (Cai, 2008). If high-dimensional models (i.e., four factors or more) are needed, estimation will be completed via the Metropolis Hastings – Robbins Monro algorithm (e.g., Cai, 2010) and SEs will be estimated via the Louis formula (Louis, 1982) with 2500 iterations.

For confirmatory item factor analysis (CIFA) models, factor loadings and overall fit of the model will be reported (Table 30). The overall fit of models examined in flexMIRT will be evaluated using the Tucker-Lewis Index (TLI, Tucker & Lewis, 1973) and the limited-information M_2-based root mean square error of approximation (RMSEA, Maydeu-Olivares et al., 2011). The cut-off for adequate fit of the TLI will be 0.97 (Cai et al., 2023) with values at or greater than 0.97 indicating good fit. For the RMSEA, the appropriate cutoff can vary based on a variety of factors, including number of categories presented for item responses. Given that CIFA analyses often involve collapsing categories to account for sparseness in responses, the varying numbers of potential item response categories make firm guidelines for the RMSEA value unclear. As a result, a cut-off value of less than 0.05 (inferred from Monroe & Cai, 2015) will be used, though failing to meet this criterion will not be considered exclusively disqualifying for the CIFA model.

Table 29. Polychoric item intercorrelations of the MiCOAS DA items using one randomly selected day from Month 1

Content	Item #	1	2	3	4	5	6	7	8	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41
Did you have a good day	1	1.00																								
Did you have a migraine day	2	0.xx	1.00																							
Did you have a headache or head pain	3	0.xx	0.xx	1.00																						
How long did the headache last	4	0.xx	0.xx	0.xx	1.00																					
Pain worse with activity	5	0.xx	0.xx	0.xx	0.xx	1.00																				
Throbbing/pounding head pain	6	0.xx	0.xx	0.xx	0.xx	0.xx	1.00																			
Pain only or worse on one side	7	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00																		
Take medication to alleviate pain	8	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00																	
Did you experience any of the following																										
Nausea	9	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00																
Vomiting	11	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00															
Sensitivity to light	13	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00														
Sensitivity to sound	15	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00													
Neck pain	17	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00												
Aura	19	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00											
Mentally slow or foggy	21	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00										
Tired	23	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00									
Were any of the following difficult?																										
Ability to think clearly	25	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00								
Ability to concentrate	27	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00							
Ability to get ready for the day	29	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00						
Ability to keep your plans	31	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00					
Ability to move your body	33	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00				
Ability to walk	35	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00			
Ability to move your head	37	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00		
Ability to bend forward	39	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00	
Ability to do strenuous physical activity	41	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	0.xx	1.00

Correlation tables will also be provided for each item subset contributing to an *a priori* measure (i.e., Symptoms, PF, SMA).

Substantively, *a priori* confirmatory models are planned for testing. For the MiCOAS DA, 3 unidimensional models, each intended to produce 1 overall score, will be examined. Path diagrams for each of the *a priori* measures from the DA items (a Symptom measure, a PF measure, and a SMA measure,) are presented in Figure 11 to Figure 13; Table 25 to Table 27 provide the full text for each of the items.

Figure 11. Path diagram for the *a priori* unidimensional model for the MiCOAS DA Symptom measure, with expected needed residual correlations among select item pairs

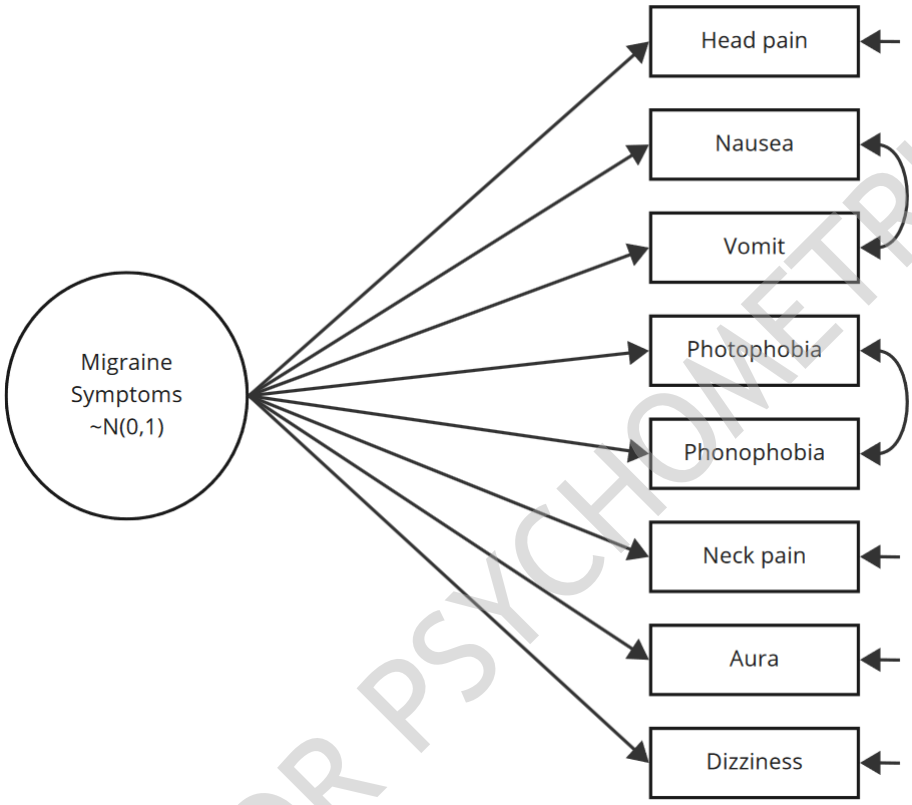


Figure 12. Path diagram for the *a priori* unidimensional model for the MiCOAS DA PF measure

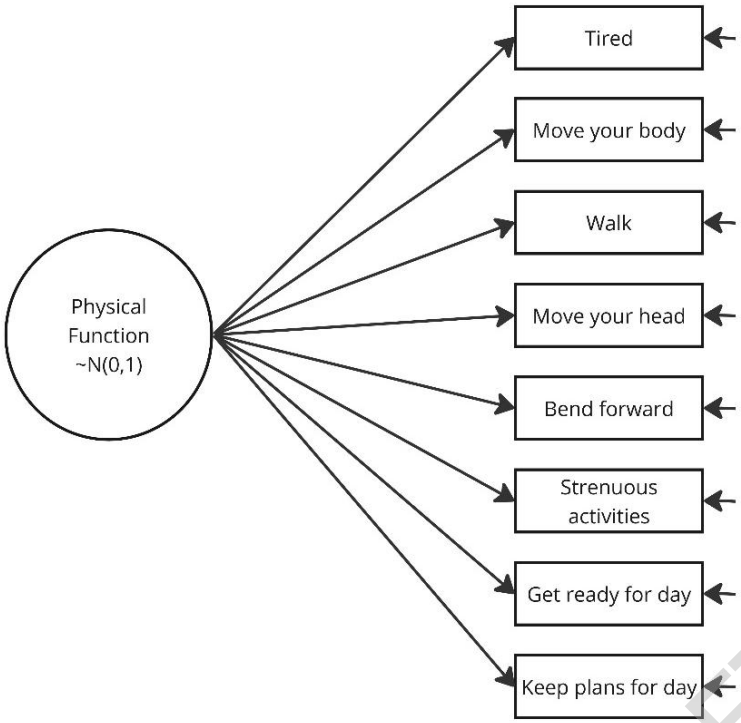
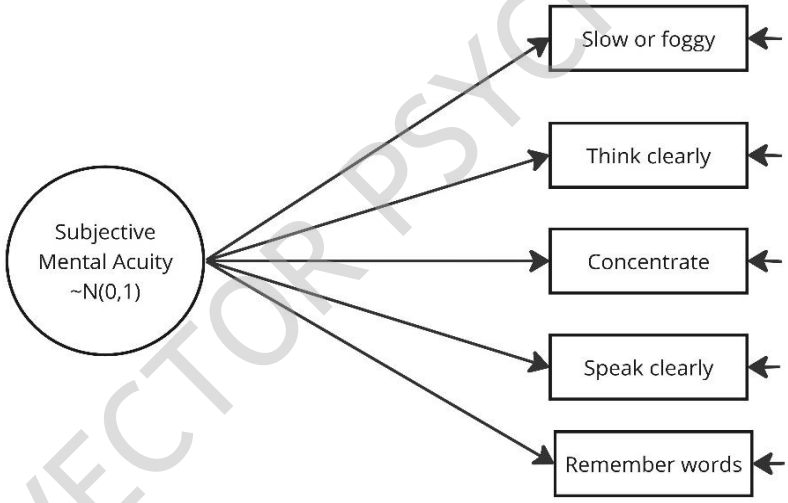


Figure 13. Path diagram for the *a priori* unidimensional model for the MiCOAS DA SMA measure



For each examined CIFA model, factor loadings and overall model fit values will be reported (Table 30).

Table 30. CIFA factor loadings and overall model fit values of the MiCOAS DA items in the PF measure

Item content	Loading
Did you have a headache or head pain	0.xx
Did you experience any of the following	
Nausea	0.xx
Vomiting	0.xx
Sensitivity to light	0.xx
Sensitivity to sound	0.xx
Neck pain	0.xx
Aura	0.xx
Dizziness	0.xx
Model fit	
RMSEA	0.xx
TLI	0.xx

Similar tables provided for the other 2 *a priori* measures from the MiCOAS DA items

As with the MiCOAS recall-based analyses, if confirmatory models for the DA measures fail to achieve adequate fit, EIFA models will be explored (and the EIFA methods detailed previously will be used for these analyses).

Table 31. EIFA factor loadings and factor intercorrelations of the MiCOAS DA Symptom items using one randomly selected day from Month 1 (if needed)

Item content	F1	…	F_last
Did you have a headache or head pain	0.xx	0.xx	0.xx
Did you experience any of the following			
Nausea	0.xx	0.xx	0.xx
Vomiting	0.xx	0.xx	0.xx
Sensitivity to light	0.xx	0.xx	0.xx
Sensitivity to sound	0.xx	0.xx	0.xx
Neck pain	0.xx	0.xx	0.xx
Aura	0.xx	0.xx	0.xx
Dizziness	0.xx	0.xx	0.xx
Factor correlations			
F1	1		
…	0.xx	1	
F_last	0.xx	0.xx	1

Similar table provided for other MiCOAS DA measures (as needed)

Item Response Theory Analyses

If a well-fitting and substantively meaningful CIFA model is found for any of the MiCOAS DA measures this model will be reparameterized within the IRT framework to examine the psychometric performance of the individual items and the measures' scores in more depth. IRT parameters will be obtained, using the categorical data analysis methods and implementations described previously. Item parameters and overall model fit and item fit statistic values (Table 32), item trace lines curves (Figure 6), and test reliability functions (Figure 7) will be reported for each measure (as available, given the dimensionality of the final model).

Table 32. IRT item parameters and item-level fit values for the MiCOAS DA Symptom measure

Item content	a/slope	b1	b2	b3	b4	S - X^2 value (df), p
Did you have a headache or head pain	x.xx	x.xx	x.xx	---	---	xx.x (xx), p = 0.xx
Did you experience any of the following						
Nausea	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Vomiting	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Sensitivity to light	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Sensitivity to sound	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Neck pain	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx
Aura	x.xx	x.xx	x.xx	x.xx	x.xx	xx.x (xx), p = 0.xx

Similar table created for other MiCOAS DA measures.

Additionally, for each DA measure, differential item functioning (DIF) analyses will be completed for groups defined by select demographic and descriptor variables. Currently, this includes chronic versus episodic migraine (CM and EM, respectively) and headache day versus non-headache day (for DA only). DIF will be implemented using the two-stage process described in Woods et al. (2013), in which a "sweep" analysis is first run to identify strongly invariant items across groups and then a more typical anchored analysis is used for the final DIF analysis, using the invariant items defined in the "sweep" step as anchor items. DIF will be tested using the Langer-improved Wald test (Langer, 2008). Overall Langer-improved Wald test p-values will be judged against Benjamini-Hochberg (1995) corrected p-values to prevent Type I errors due to multiple testing. If the overall DIF test for an item is found to be statistically significantly different across groups, examinations into the meaningfulness/severity of the detected DIF (e.g., Hansen et al., 2014) using the weighted area between the curve (wABC) to quantify differences between groups will be completed prior to final determinations regarding flagging an item for non-trivial DIF. Items with a wABC value greater than 0.30 (Hansen et al., 2014) will be considered to exhibit non-ignorable DIF.

Reliability

Reliability is a measure of the extent of consistency or repeatability of scores obtained from a given model. The target PRO measures will be assessed for both internal consistency and test-retest reliability.

Internal Consistency Reliability

Internal consistency reliability evaluates individual items in comparison with one another for their ability to give consistently appropriate results. To investigate internal consistency of the items contributing to the MiCOAS DA measures separately (the 3 *a priori* DA measures, assuming CIFA results are supportive), CTT analyses, (i.e., coefficient alpha, alpha with item *i* removed, and item-total correlations) will be reported using data from the previously described randomly selected day from Month 1 dataset and values will be reported similar to Table 33.

If a multidimensional solution is found as the preferred model for any of the target PRO measures, the items contributing to each dimension will be submitted to CTT analysis separately. Additionally, the marginal (distributional) reliability value from the IRT-based analyses will be reported. A value of 0.70 will be used as the minimum overall coefficient alpha and the minimum marginal reliability value to demonstrate satisfactory internal consistency (e.g., Nunnally, 1978).

Table 33. Coefficient alpha, item-total correlations, alpha-if-deleted for MiCOAS DA Symptoms items in randomly selected Month 1 dataset

Parameter	Value	
N	xxx	
Overall alpha	0.xx	
Item	Item-total correlation	Alpha-if-deleted
Did you have a headache or head pain	0.xx	0.xx
Did you experience any of the following		
Nausea	0.xx	0.xx
Vomiting	0.xx	0.xx
Sensitivity to light	0.xx	0.xx
Sensitivity to sound	0.xx	0.xx
Neck pain	0.xx	0.xx
Aura	0.xx	0.xx

Test-retest reliability

Test-retest reliability will be assessed for the each of the MiCOAS measures scores (Table 14). Test-retest will be evaluated using uncorrected Pearson correlations as well as intraclass correlation coefficients (ICCs) from a two-way mixed effect model with absolute agreement for single measures (McGraw & Wong, 1996). ICCs

will be judged against thresholds established for interpreting intraclass correlations (i.e., values less than 0.40 indicate "poor," 0.40 to 0.75 as "fair to good;" 0.75 to 0.90 as "good," [Fleiss, 1986]).

For the MiCOAS DA measures, test-retest will be examined for 3 separate summary score derived from the DA data. First, in Cohort 3, weekly summary score (found as the average of the 7 days contributing to a given week) will be examined between Week 1 and Week 2. Second, biweekly scores will be investigated in Cohorts 1 and 2, in which average DA summary scores over 14 days (2 weeks) will used, specifically between Week 2 and Week 4. Additionally, to mirror the treatment of DA data seen in clinical trial analyses, a 28-day average will be calculated for each measure score and test-retest will use Month 1 (that is, average of Day 1 to Day 28) and Month 2 (average of Day 29 to Day 56) values from Cohort 1. Summary scores over a given time period (weekly, biweekly, monthly) will only be calculated for observations with non-missing DA entries on 50% or more of the days in the time period (i.e., 4 days required for a weekly score to be calculated, 7 days required for biweekly scores, 14 days required for monthly scores).

Because the variability and change often seen in observational migraine studies as well as in placebo groups of randomized clinical trials, test-retest analyses will be performed on a "stable" subset of participants. For the "random day" test-retest, stable participants will be defined as those whose PGIS did not change between the end of the first time period and the end of the second time period; as a example, for the weekly test-retest, the change in the PGIS administered at Week 1 (Day 7) and the PGIS administered at Week 2 (Day 14). If an insufficient number of stable participants (fewer than 30) are identified by these definitions, results will be reported with a caution regarding overinterpretation of values derived from small samples.

Table 34. Test-retest values for MiCOAS DA measure scores using a stable subset of participants.

Time period / Measure	Pearson correlation	ICC
Weekly (Week 1 [Day 1 to 7] average of DA to Week 2 [Day 8 to 14] average of DA)		
Symptoms	0.xx	0.xx
PF	0.xx	0.xx
SMA	0.xx	0.xx
Biweekly (Week 2 [Day 1 to 14] average of DA to Week 4 [Day 15 to 28] average of DA)		
Symptoms	0.xx	0.xx
PF	0.xx	0.xx
SMA	0.xx	0.xx
Monthly (Month 1 [Day 1 to 28] average of DA to Month 2 [Day 29 to 56] average of DA)		
Symptoms	0.xx	0.xx
PF	0.xx	0.xx
SMA	0.xx	0.xx

VALIDATION ANALYSES

Several relationships with other assessments will be examined using MiCOAS DA scores. First, in Cohort 3, weekly average scores (found as the average of the 7 days contributing to a given week) will be examined. Second, biweekly average scores will be investigated in Cohorts 1 and 2, in which average DA summary scores over 14 days (2 weeks) will used. Finally in Cohort 1, monthly scores (using a 28-day average) will be calculated for each measure score. Average DA scores over a given time period (weekly, biweekly, monthly) will only be calculated for observations with non-missing DA entries on 50% or more of the days in the time period (i.e., 4 days required for a weekly score to be calculated, 7 days required for biweekly scores, 14 days required for monthly scores).

Given the exploratory nature of these analyses, no family-wise error rate corrections will be made to account for multiple testing. Additionally, the number and interpretation of scores from the MiCOAS DA item sets will depend on the results of the factor analyses; the expectations stated below are based on the assumption that the *a priori* unidimensional models of 3 measures are supported and will produce individual scores (e.g., ~~Symptom~~MSI, PF, and SMA scores from the DA items).

The first validity evidence to be presented will be descriptive statistics of each MiCOAS DA average score at the target timepoints. Available sample size, mean, and SD for each score will be reported for each DA score at each target timepoints (e.g., Week 1, 2, 3, and 4 for weekly scores in Cohort 3; Table 35).

Table 35. Descriptive statistics of the weekly MiCOAS DA scores at target timepoints

Timepoint/ Descriptive	Symptoms	PF	SMA
Week 1 (Day 7)			
N	xx	xx	xx
M (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx
Week 2 (Day 14)			
N	xx	xx	xx
M (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx
Week 3 (Day 21)			
N	xx	xx	xx
M (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx
Week 4 (Day 28)			
N	xx	xx	xx
M (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx	xx	xx
Min, Max	xx, xx	xx, xx	xx, xx

Similar tables presented for biweekly and monthly average scores at relevant timepoints

The expected direction between the *a priori* measure scores of the MiCOAS DA and reference variables are specified in Table 36 and observed values will be reported within each target timepoint (e.g., Table 37,).

Table 36. Expected direction and magnitude among the MiCOAS DA measure scores and reference variables

Variable	Symptoms	PF	SMA
Age	-0.2 to 0.2	0 to 0.3	0 to 0.3
Height	-0.2 to 0.2	-0.2 to 0.2	-0.2 to 0.2
MOS-Cog-R	-0.3 to 0	-0.3 to 0	-0.5 to -0.3
HIT-6 Total score	0.4 or greater	0.4 or greater	0.4 or greater
Item: Limit ability to concentrate	0.1 to 0.3	0.1 to 0.3	0.4 or greater
Item: Too tired to do work	0.1 to 0.3	0.3 to 0.5	0.4 or greater
MFIQ v 2.0			
Physical function	0.3 to 0.5	0.4 or greater	0.3 to 0.5
Usual activities	0.3 to 0.5	0.4 or greater	0.3 to 0.5
Social function	0.1 to 0.3	0.1 to 0.3	0.1 to 0.3
Emotional function	0.1 to 0.3	0.1 to 0.3	0.1 to 0.3
Item: Difficult...activities required concentration	0.1 to 0.3	0.1 to 0.3	0.3 to 0.5
MSQ v2.1			
Role function - restrictive	-0.3 to 0	-0.5 to -0.3	-0.3 to 0
Role function - preventive	-0.3 to 0	-0.5 to -0.3	-0.3 to 0
Emotional function	-0.3 to 0	-0.3 to 0	-0.3 to 0
Item: limit ability to concentrate	-0.3 to 0	-0.3 to 0	-0.5 or more extreme
MMD†	0.3 to 0.5	0.2 to 0.5	0.2 to 0.5

Note. PF = Physical functioning. SMA = Subjective mental acuity. † = only available at Week 4 and Week 8.

Table 37. Correlations among the weekly MiCOAS DA measures and reference variables at Week 1, 2, and 3

	Symptoms		PF		SMA	
Reference measure	N	r	N	r	N	r
MOS-Cog-R 7d	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MFIQ v2 (7-day recall)						
Physical Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Usual Activities	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Social Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***

Note. PF = Physical functioning. SMA = Subjective mental acuity.

Table 38. Correlations among the weekly MiCOAS DA measures and reference variables at Week 4

	Symptoms		PF		SMA	
Reference measure	N	r	N	r	N	r
MOS-Cog-R 7d	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
HIT-6 (4-week recall)	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Too tired to do work	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MSQ v2.1 (4-week recall)						
Role Function - Restrictive	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Role Function- Preventative	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Difficult...activities required concentration	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MFIQ v2 (7-day recall)						
Physical Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Usual Activities	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Social Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MMD	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***

Note. PF = Physical functioning. SMA = Subjective mental acuity.

Tables similar to Table 37 or Table 38 provided for biweekly and monthly scores are relevant timepoints

Known-Groups Evidence

Known-groups analyses are examined to determine the extent to which the MICOAS average measure scores reflect expected differences between clinically meaningful and conceptually distinct groups.

For the target PRO measures known-groups analyses, clinically meaningful groups will be formed from the following reference variables:

- PGIS response categories: it is expected that as PGIS categories become more severe (from None to Very Severe) the weekly/biweekly/monthly MiCOAS DA scores will also increase.
- PGIC response categories (only examined at Week 4 and Week 8): it is expected that as PGIC response categories increase (from Very Much Worse to Very Much Improved) the weekly/biweekly/monthly MiCOAS DA scores will also increase.
- HIT-6 groups: it is expected that as HIT-6 severity category increases (from Little to No Impact up through Severe impact), the weekly/biweekly/monthly MiCOAS DA will also increase.

For the scores from the weekly/biweekly/monthly MiCOAS DA measures, cross-sectional known-groups analyses within each target timepoint are planned (e.g., Table 18). For these continuous scores, group sample

size, means, and SDs will be reported and ANOVA will be used to examine differences in each measure's scores across clinically meaningful groups. For all known-groups analyses, no post-hoc tests to probe statistically significant overall ANOVA effects will be conducted and, to prevent the over-interpretation of results derived from small n values, analysis results will only be reported if groups constructed from the reference variables have at least 25 observations.

Table 39. Known-groups analyses of weekly MiCOAS DA PF measure scores by timepoint

	Week 1			Week 2			Week 3			Week 4		
Anchor/Groups	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
PGIS	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
None	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Mile	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Moderate	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Severe	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Very severe	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Overall Effect (ANOVA)	F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx		
PGIC												
Very much improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Much improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Minimally improved				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
No change				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Minimally worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Much worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Very much worse				xxx	x.x	x.x	xxx	x.x	x.x	xxx	x.x	x.x
Overall Effect (ANOVA)				F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx			F(x,xxx)=xx.x, p=0.xx		
HIT-6 groups												
Little to no impact										xxx	x.x	x.x
Moderate impact										xxx	x.x	x.x
Substantial impact										xxx	x.x	x.x
Severe impact										xxx	x.x	x.x
Overall Effect (ANOVA)										F(x,xxx)=xx.x, p=0.xx		

Similar tables provided for other measures from the weekly average MiCOAS DA scores and for biweekly and monthly average scores at relevant timepoints

Ability to Detect Change

The ability to detect change over time is critical for understanding the adequacy of measures for clinical research. In short, if patients have changed over the course of the study, that change will be reflected in the values of scores that are valid for their intended purpose. Given the observational nature of the quantitative study, it is unknown what level of change over time will be realized in the data, but in consideration of the fact that migraine observational studies and placebo groups in migraine trials do tend to change, the ability of detect change of each of the MiCOAS measures' scores will be investigated.

The assessment of ability to detect change will be examined by creating change scores for the weekly/biweekly/monthly MiCOAS DA which will be calculated as the average DA score over time period 1 value minus the time period 2 values so that positive change scores indicate improvement. For the weekly scores, change will be defined as Week 1 – Week 4 in Cohort 3 only. For the biweekly scores in Cohort 1 and 2, change will be defined as Week 2 – Week 4. For the monthly scores in Cohort 1 only, change will be defined as Week 4/Month 1 – Week 8/Month2 average values. Correlations among the change in reference variables and the change in each weekly/biweekly/monthly MiCOAS DA score will be examined. The direction and magnitude of the specified correlations are expected to mirror the relationships described in the convergent and discriminant evidence section in Table 36 and will be reported in a format similar to Table 40.

Table 40. Correlations of change in weekly MiCOAS DA scores with change in reference variables (Week 1 – Week 4)

	Symptoms		PF		SMA	
Change in reference measure	N	r	N	r	N	r
MOS-Cog-R 7d	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
HIT-6 (4-week recall)	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Too tired to do work	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MSQ v2.1 (4-week recall)						
Role Function - Restrictive	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Role Function- Preventative	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: Difficult...activities required concentration	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
MFIQ v2 (7-day recall)						
Physical Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Usual Activities	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Social Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Emotional Function	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***
Item: limit ability to concentrate	xxx	0.xx ***	xxx	0.xx ***	xxx	0.xx ***

Similar table provided for biweekly change scores and monthly change scores. Note that monthly change score correlations will also include change in MMD as a reference variable.

MEANINGFUL SCORE REGION ANALYSES

Meaningful score regions (MSRs) will be estimated for weekly/biweekly/monthly MiCOAS DA score using Week 1/Week 2/Week 4 average DA scores. This analysis will employ an anchor-based approach using responses on the PGIS to identify MSR for the each of the weekly/biweekly/monthly MiCOAS DA average scores. Using the responses of the PGIS item at Week 1/Week 2/Week 4, summary values that describe the distribution of the weekly/biweekly/monthly MiCOAS summary scores within each response group will be constructed (Table 41), along with boxplots to visually depict these distributions (top section of Figure 8).

Table 41. Summary of weekly MiCOAS DA scores by PGIS item responses at Week 1 (Day 7)

PGIS at Week 1 (Day 7)	n	Min	Q1	Mean	Median	Q3	Max	SD
			Symptoms					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			PF					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
			SMA					
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity. PF = Physical functioning. SMA = Subjective mental acuity.

Similar tables provided for biweekly and monthly scores using appropriate PGIS

To define the cut-values between response categories (based on FDA suggestion for such requests in other projects), the 75th percentile (Q3) of each MiCOAS summary score of the lower response category is used to define the cut-value between two PGIS response categories. For example, if the “None” PGIS response is found to have a range of a given MICOAS score (which is planned to be on a t-distribution with a mean of 50 and a SD of 10) from 25 to 41 with a score of 34 found to be in the 75th percentile of the distribution of scores, the value of 34 is used to define the threshold to separate the “None” from “Mild” score ranges. A similar process will be used to define the cut-values between subsequent PGIS response categories, resulting in non-overlapping categories of each MiCOAS score for groups defined by PGIS ratings of severity, which can eventually be used to assist in the interpretation of observed differences in treatment group scores (bottom section of Figure 8).

MEANINGFUL WITHIN-PERSON CHANGE/MSD THRESHOLD ESTIMATES

As noted earlier, it is unknown what level of change over time will be realized in the data, but in consideration of the fact that migraine observational studies and placebo groups in migraine trials do tend to change and the importance of thresholds to aid in the interpretation of scores, exploratory analyses to begin to investigate meaningful within-person change threshold value for each of the MiCOAS DA average scores will be investigated.

A MSD_w threshold sets the minimum amount of individual change necessary for a patient to experience a "meaningful" change in their condition over time; such thresholds are important tools for understanding and interpreting the observed change in an outcome of interest. For the MiCOAS DA average scores, improvement and worsening MSD_w thresholds will be estimated (if change over time in a sufficient number of participants is observed). Both anchor- and distribution-based methods will be used to investigate change in the weekly/biweekly/monthly MiCOAS DA scores and obtain candidate MSD_W threshold values for each of the target migraine PRO measure scores (if possible) for each of the noted time period average scores.

Anchor-based methods

To establish the suitability of potential anchors, correlations among change scores and the candidate anchor variables (change in the PGIS, PGIC responses at relevant time, change in the HIT-6 total score) will be examined against the change in each of the MiCOAS DA average scores.

The assessment of ability to detect change will be examined by creating change scores for the weekly/biweekly/monthly average MiCOAS DA, which will be calculated as the average DA score over time period 1 value minus the time period 2 values so that positive change scores indicate improvement. For the weekly scores, change is defined as Week 1 – Week 4 in Cohort 3 only. For the biweekly scores in Cohort 1 and 2, change is defined as Week 2 – Week 4. For the monthly scores in Cohort 1 only, change is defined as Week 4/Month 1 – Week 8/Month2 average values.

A priori, the change in the PGIS item is considered the primary anchor for the weekly/biweekly/monthly MiCOAS DA scores and meaningful change is defined as an improvement or worsening of 1 response category (e.g., Severe to Moderate for improvement, Moderate to Severe for worsening) between the first and second timepoint. For the PGIC, a response of "Minimally better" is considered meaningful improvement and a response of a "Minimally worse" is considered meaningful worsening. For the change in the HIT-6, a decrease of 6 is considered meaningful improvement (Houts et al., 2020); there is no empirically derived worsening threshold for the HIT-6 total score, but an increase of 6 (consistent with the improvement threshold) will be considered meaningful worsening.

For candidate anchors to be suitable for use, it is generally desirable that the weekly/biweekly/monthly average MiCOAS DA change scores be correlated with the potential anchor at least 0.3 (e.g., Coon & Cook,

2018; Revicki et al., 2008). Spearman correlations between the change in the weekly/biweekly/monthly MiCOAS DA scores and change in the potential anchors will be reported to allow for a determination of suitability of use as an anchor with each weekly/biweekly/monthly MiCOAS DA score.

To examine the extent to which the observed change over time is influenced by baseline status, a table of key percentiles of the target COA change scores, stratified by participants baseline status on the *a priori* primary anchor (change in the PGIS) will be constructed (Table 21), similar to FDA's patient-focused drug development guidance 4's Table 1 (pg. 23).

Table 42. Key percentiles of MiCOAS DA Symptoms change scores by Week 1 (Day 1) PGIS response for all Cohort 3 participants

		Change in MiCOAS score, by key percentile				
PGIS at Week 1 (Day 7)	n (%)	10th	25th	50th	75th	90th
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity.

Similar tables constructed for all other weekly MiCOAS DA measures and all biweekly and monthly MiCOAS DA scores.

Additionally, to understand the change in those participants who meaningfully improved or worsened, a similar table of key percentiles for only those participants who meaningfully changed (Table 43) will be reported.

Table 43. Key percentiles of weekly MiCOAS DA Symptoms change scores by Week 1 (Day 7) PGIS response for Cohort 3 participants who improved or worsened 1-category on the PGIS between Week 1 (Day 7) and Week 4 (Day 28)

		Change in MiCOAS score, by key percentile				
PGIS at Week 1 (Day 7)	n (%)	10th	25th	50th	75th	90th
		Meaningfully Improved				
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Very severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
		Meaningfully Worsened				
None	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Mild	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Moderate	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx
Severe	xxx (xx.xx%)	xx.xx	xx.xx	xx.xx	xx.xx	xx.xx

Note. n = Group sample size. % = Percent. PGIS = Patient Global Impression of Severity. PGIS responses of None in the meaningfully improved section and Very severe in the Meaningulkly worsened sections have been excluded as it is not possible to improve/worsen from these responses.

Similar tables constructed for all other weekly MiCOAS DA measures and all biweekly and monthly MiCOAS DA scores.

For the anchors that are acceptable for use with each of the weekly/biweekly/monthly MiCOAS DA scores, all categories will be used to define groups for determining candidate MSD_w values for improvement. Additionally, to ensure that illogical threshold values are not chosen (e.g., for the PGIC item, the “No change” group having a smaller change than the “Minimally improved” group), descriptive statistics of the weekly/biweekly/monthly MiCOAS DA change scores at each level of each anchor item will be reported (e.g., Table 44).

Table 44. Descriptive statistics of weekly MiCOAS DA average change scores (Week 1 [Day 7] – Week 4 [Day 28]) by anchor variables

	Symptoms					PF					SMA				
Anchor variable/Levels of anchor variable	n	M	Median	SD	Min, Max	n	M	Median	SD	Min, Max	n	M	Median	SD	Min, Max
Change in PGIS response															
3+ category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
2 category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
1 category improvement	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
1 category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
2 category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
3+ category worsening	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
PGIC response															
Very much improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Much improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Minimally improved	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Minimally worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Much worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Very much worse	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Change in HIT-6															
Improvement responder (+6 or more better)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
No change (-6 to 6)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx
Worsening responder (-6 or more worse)	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx	xxx	x.xx	x.xx	x.xx	xx, xx

Note. n = group size. M = Mean. SD = Standrd deviation. Min = minimum. Max = Maximum. PF = Physical function. SMA = Subjective mental acuity.

Similar tables constructed for all biweekly and monthly MiCOAS DA average change scores.

The approach for identifying thresholds for interpretation will be the presentation of empirical cumulative distribution functions (eCDFs) and empirical probability distribution functions (ePDFs). eCDFs plot the cumulative proportion of subjects reporting change scores on a COA along the range of the COA change scores on the x-axis. eCDFs will present separate curves for each level of change on the anchor (e.g., Figure 9). While there is no universal guideline for interpreting such eCDFs, the location at which the meaningful deterioration curve reaches 50% on the y-axis is a possible location for an interpretable threshold. The ePDF curves (e.g., Figure 10) are useful in aiding the interpretation of eCDF curves. While presenting the same information as the eCDFs, ePDF curves may provide a more interpretable overview of the shape, dispersion, and skewness of the distribution of the change from baseline in the endpoint of interest across various anchor categories.

Distribution-based methods

For the distribution-based estimates, 2 typical values will be reported for each of the weekly/biweekly/monthly MiCOAS DA measure scores: 1) ½ SD of timepoint 1 scores and 2) the standard error of measurement (SEM) at timepoint1. For the weekly scores, timepoint 1 is Week 1 (Day 7), for the biweekly scores, timepoint 1 is Week 2 (Day 14), and for the monthly scores, timepoint 1 is Week 4 (Day 28). The previously found internal consistency alpha value for the DA items will be used as the reliability estimate in calculating the SEMs, using the formula $\mathrm{SEM} = \mathrm{SD} * \sqrt{(1 - \mathrm{reliability})}$. While these distribution-based values are not patient-centered, they will be used to set a minimum value that any anchor-based suggested MSD_W threshold value must exceed to ensure that a change over time that could solely be attributable to measurement error/noise in the weekly/biweekly/monthly average MiCOAS DA scores is not considered "meaningful".

Triangulation

The candidate meaningful MSD values based on the anchor- and distribution-based methods described above will be collected into a table for improvement (e.g., Table 24) and, separately for worsening. Tabled values will be considered to identify a value or a range of values that may be used for interpreting change scores (i.e., triangulation). Distribution-based values, as noted above, will be used to set a minimum change value, while results from the anchor-based methods will be evaluated and weighted with consideration of the content relevance and strength of correlation of the anchor with the weekly/biweekly/monthly average MiCOAS DA scores during the triangulation process to establish a final MSD_W value/range of values.

Table 45. Candidate anchor- and distribution-based improvement MSD values for weekly/biweekly/monthly average MiCOAS DA change scores

	Anchor-based			Distribution-based (at Timepoint 1)	
	Change in PGIS	PGIC	HIT-6 Total		
Timeperiod / Measure	+1 Category Median	Minimally better Median	Improved Median	1/2 SD	SEM
Weekly					
Symptoms	x.xx	x.xx	x.xx	x.xx	x.xx
PF	x.xx	x.xx	x.xx	x.xx	x.xx
SMA	x.xx	x.xx	x.xx	x.xx	x.xx
Biweekly					
Symptoms	x.xx	x.xx	x.xx	x.xx	x.xx
PF	x.xx	x.xx	x.xx	x.xx	x.xx
SMA	x.xx	x.xx	x.xx	x.xx	x.xx
Monthly					
Symptoms	x.xx	x.xx	x.xx	x.xx	x.xx
PF	x.xx	x.xx	x.xx	x.xx	x.xx
SMA	x.xx	x.xx	x.xx	x.xx	x.xx

Note. PF = Physical functioning. SMA = Subjective mental acuity. PGIS = Patient Global Impression of Severity, PGIC = Patient Global Impression of Change.

Note. Similar table provided for worsening (provided sufficient data).

INTERPRETATION AND CONCLUSIONS

"Validity" is not a one-time "Yes or No" decision, but rather an on-going accumulation of evidence over numerous studies using the instrument/item to reach a conclusion regarding the degree to which this body of work supports the interpretations intended when using its scores in a given context (e.g., Edwards et al., 2018). In consideration of this, the pattern of results across all validity analyses will be examined for each MiCOAS migraine measure and a holistic account of the degree to which current evidence regarding the suitability of that item for use to support valid inferences in the current context of use will be made. It is expected that the majority of observed correlations will conform to expectations with respect to direction and relative strength, that the majority of values from known-groups analyses will demonstrate non-trivial differences (in the expected direction) between clinically meaningful groups, and that the measure scores of the MiCOAS migraine questionnaires will be shown to be sufficiently sensitive to detect known change.

REFERENCES

Bayliss, M.S. & Batenhorst, A.S. (2002). The HIT-6™: A User's Guide. Lincoln, RI: QualityMetric Incorporated.

Benjamini, Y., & Hochberg, Y., (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.

Browne, M. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 61, 309–329.

Cai, L. (2010). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35, 307–335.

Cai, L. (2024). flexMIRT® version 3.7: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Cai, L., Chung, S. W., & Lee, T. (2023). Incremental model fit assessment in the case of categorical data: Tucker–Lewis index for item response theory modeling. *Prevention Science*, *24*(3), 455-466.

Coon, C.D., & Cook, K.F. (2018). Moving from significance to real-world meaning: Methods for interpreting change in clinical outcome assessment scores. *Quality of Life Research*, 27, 33-40.

Edwards, M.C., Slagle, A., Rubright J.D., & Wirth, R.J. (2018). Fit for purpose and modern validity theory in clinical outcomes assessment. *Quality of Life Research*, 27(7), 1711-1720.

Fleiss, J.L. (1986). Reliability of measurement. In: *The Design and Analysis of Clinical Experiments* [pp. 1-32]. New York: Wiley.

Hansen, M, Cai, L., Stucky, B.D., Tucker, J.S., Shadel, W.G., & Edelen, M.O. (2014). Methodology for developing and evaluating the PROMIS® smoking item banks. *Nicotine and Tobacco Research*, 16(3), S175-S189.

Houts, C.R., Wirth, R.J., McGinley, J.S., Cady, R. & Lipton, R.B. (2020). Determining thresholds for meaningful change for the Headache Impact Test (HIT-6) total and item-specific scores in chronic migraine. *Headache*, 60(9), 2003-2013.

Kawata, A.K., Hareendran, A., Shaffer, S., Mannix, S., Thach, A., Desai, P., Mikol, D.D., Ortmeier, B., Bayliss, M., & Buse, D.C. (2019). Evaluating the psychometric properties of the Migraine Functional Impact Questionnaire (MFIQ). *Headache, 59*, 1253 – 1269.

Kosinski, M., Bayliss, M. S., Bjorner, J. B., Ware Jr, J. E., Garber, W. H., Batenhorst, A., ... & Tepper, S. (2003). A six-item short-form survey for measuring headache impact: The HIT-6™. *Quality of Life Research*, *12*(8), 963-974.

Langer, M. (2008). A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.

Liu, O.L., & Rijmen, F. (2008). A modified procedure for parallel analysis of ordered categorical data. *Behavioral Research Methods*, 40(2), 556-562.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 44, 226–233.

Martin, B.C., Pathak, D.S., Sharfman, M.I., Adelan, J.U., Taylor, F., Kwong, J., & Jihngran, P. (2000). Validity and reliability of the Migraine-specific Quality of Life Questionnaire (MSQ version 2.1). *Headache*, 40, 204-215.

Maydeu-Olivares, A., Cai, L., & Hernandez, A. (2011). Comparing the fit of IRT and factor analysis models. *Structural Equation Modeling*, 18, 333–356.

McGraw K.O., & Wong S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46.

Monroe, S. & Cai, L. (2015). Evaluating structural equation models for categorical outcomes: A new test statistic and a practical challenge of interpretation. *Multivariate Behavioral Research*, 50(6), 569-583.

Nunnally, J. C. (1978). *Psychometric Theory*, 2nd ed. New York: McGraw-Hill.

Revicki, D., Hays, R.D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61, 102-109.

Tucker, L. R, & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.

Wirth, R.J., & Edwards, M.C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12, 58-79.

Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*, *73*(3), 532-547.

Yarlas, A., White, M.K., Bjorner, J.B. (2013). The development and validation of a revised version of the Medical Outcomes Study Cognitive Functioning Scale (MOS-COG-R). *Value in Health*, 16, A33.