Measurement properties of instruments for migraine impact in Brazil: a systematic review

Medida de instrumentos para impacto da enxaqueca no Brasil: uma revisão sistemática

Medición de instrumentos para el impacto de la migraña en Brasil: una revisión sistemática

Mayara Cristina Macêdo de Menezes
ORCID: https://orcid.org/0000-0002-0239-2234
Universidade Federal de Pernambuco, Brazil
E-mail: mayaracmm@gmail.com
Andrea Lemos Bezerra de Oliveira
ORCID: https://orcid.org/0000-0003-0631-0512
Universidade Federal de Pernambuco, Brazil
E-mail: andrealemos4@gmail.com
Hugo Gabriel Feitosa de Souza
ORCID: https://orcid.org/0000-0003-3247-5310
Universidade Federal de Pernambuco, Brazil
E-mail: hugogfeitosa@gmail.com
Renata Crespo Simas Toscano
ORCID: https://orcid.org/0000-0002-1015-7816
Universidade Federal de Pernambuco, Brazil
E-mail: renatinhacrespo@gmail.com
Daniella Araújo de Oliveira
ORCID: https://orcid.org/0000-0002-6013-978X
Universidade Federal de Pernambuco, Brazil
E-mail:sabinodaniellaufpe@gmail.com

Abstract
Objective: To critically appraise, compare and summarize the quality of the measurement properties of all *Patient-Reported Outcome Measures* translated and validated for the Brazilian population that assess the impact of migraine. Background The evaluation of measurement properties of *Patient-Reported Outcome Measures* is needed for professionals and researchers to select instruments that ensure quality of results. Thus, reliable instruments are important to provide information on the impact of migraine. Methods: The search was conducted in MEDLINE/Pubmed, Web of Science, LILACS, and Embase databases. We included studies evaluating measurement properties of PROMs that assessed the impact of migraine and were validated for Brazil. Methodological quality, risk of bias, and quality of evidence were assessed following the Consensus-based Standards for Selecting Health Measurement Instruments for systematic reviews of *Patient-Reported Outcome Measures* Results: A total of 112 studies identified, and four were included. Three instruments were analyzed: Headache Impact Test, which presented a serious risk of bias with moderate quality of evidence; Pediatric Migraine Disability Assessment, which presented an extremely serious risk of bias and very low quality of evidence; and the Headache Disability Inventory¸ which presented a very serious risk of bias and low quality of evidence. Conclusion: The three instruments evaluated had important limitations regarding quality of evidence. The Headache Impact Test was the most recommended instrument because it presented a moderate quality of evidence.
Keywords: Migraine disorders; Brazil; Patient reported outcome measures; Sickness impact profile; Teaching.

Resumo
Objetivo: Realizar uma revisão sistemática da qualidade das propriedades de medida de todas as medidas de desfecho relatadas pelo paciente (PROMs) validadas para o Brasil e desenvolvidas para avaliar o impacto da enxaqueca. Introdução: A avaliação das propriedades de medida das Medidas de Desfecho Relatado pelo Paciente é necessária para que profissionais e pesquisadores selecionem instrumentos que garantam a qualidade dos resultados. Assim, instrumentos confiáveis são importantes para fornecer informações sobre o impacto da enxaqueca. Métodos: Foi realizada uma busca nas bases de dados MEDLINE/Pubmed, Web of Science, LILACS e Embase, incluindo estudos que avaliassem propriedades de medidas de PROMs desenvolvidos para avaliação do impacto causado pela enxaqueca, e que traduzidos e validados para a população brasileira. A qualidade metodológica, risco de viés e qualidade da evidência foram avaliadas seguindo as diretrizes do *COnsensus-based Standards for the selection of health Measurement INstruments* para revisões sistemáticas de PROMs. Resultados: Um total de 112 estudos identificados, e quatro foram incluídos. Foram analisados três instrumentos: *Headache Impact* Test, que apresentou sério risco de viés com qualidade de evidência moderada; *Pediatric Migraine Disability* Assessment, que apresentou

um risco extremamente grave de viés e qualidade de evidência muito baixa; e o *Headache Disability Inventory*¸ que apresentou um risco muito sério de viés e baixa qualidade de evidência.Conclusão: A análise realizada identificou que os três instrumentos avaliados apresentaram limitações importantes na qualidade de evidência dos instrumentos avaliados. O *Headache Impact Test*, foi o mais recomendado por apresentar moderada qualidade de evidência.
Palavras-chave: Transtornos de enxaqueca; Brasil; Medidas de resultados relatadas por pacientes; Perfil de impacto da doença; Ensino.

Resumen
Objetivo: Valorar críticamente, comparar y resumir la calidad de las propiedades de medición de todas las medidas de resultados informados por el paciente, traducidas y validadas para la población brasileña que evalúa el impacto de la migraña. Introducción: La evaluación de las propiedades de medición de las medidas de resultado informadas por el paciente es necesaria para que los profesionales e investigadores seleccionen instrumentos que aseguren la calidad de los resultados. Por lo tanto, los instrumentos fiables son importantes para proporcionar información sobre el impacto de la migraña. Metodología: La búsqueda se realizó en las bases de datos MEDLINE/Pubmed, Web of Science, LILACS y Embase. Se incluyó estudios que evaluaron propiedades de medición de los instrumentos desarrollados para evaluar el impacto de la migraña traducidos y validados para la población brasileña. La calidad metodológica, el riesgo de sesgo y la calidad de la evidencia se evaluaron siguiendo el Consensus-based Standards for the selection of health Measurement Instruments para revisiones sistemáticas de las mediciones informadas por el paciente. Resultados: Se identificaron 112 estudios y se incluyeron cuatro. Se analizaron tres instrumentos: Headache Impact, lo cual presentó un riesgo de sesgo grave con calidad de evidencia moderada; Pediatric Migraine Disability Assessment que presentó riesgo de sesgo extremadamente grave y con una calidad de evidencia muy baja; por fin Headache Disability Inventory presentó un riesgo de sesgo muy serio y con una baja calidad de evidencia. Conclusión: Los tres instrumentos evaluados presentaron importantes limitaciones en cuanto a la calidad de evidencia. El Headache Impact Test fue el instrumento más recomendado por presentar una calidad de evidencia moderada.
Palabras clave: Transtornos migrañosos; Brasil; Medidas de resultados percibidos por los pacientes; Perfil de Impacto de enfermedad; Enseñanza.

1. Introduction

Migraine is considered a neurobiological disease (Goadsby et al., 2017) and identified as a preeminent cause of disability, especially in individuals of working age (Agosti, 2018). Among all the neurological diseases that cause disability, migraine occupies the 15th place, affecting 14.4% of the adult world population (Steiner et al., 2018; Steiner et al., 2020) and 16% of the Brazilian population (Queiroz; Silva Junior, 2015). The functional decline caused by migraine is similar to depression, diabetes mellitus, and myocardial infarction, and the frequency of migraine crises is associated with a high rate of disability(Agosti, 2018).

Biopsychosocial factors should be considered in people with migraine since it provides a specific, targeted, and patient-centered assessment (Edwards et al., 2006). Also, information on the frequency of migraine-attributed disability complements the medical diagnosis and helps develop the treatment plan. Therefore, reliable instruments are needed to provide global assessment measures for people with migraine and identify the impact caused by migraine in this population (Stewart et al., 2000).

Assessment instruments are used during clinical evaluation to monitor and describe neurological disorders and reduce the subjectivity of information collected from the patient. This process also aims to verify the effectiveness of the treatment through reassessment (Yang et al., 2011). Given the importance of these assessment tools in complementing clinical decisions, reliability is essential to avoid clinical decisions with inadequate conclusions(Guyatt et al., 2011; Mokkink et al., 2019).

The quality and reliability of an instrument are related to measurement properties, which must be evaluated to classify the content validity of a patient-reported outcome measure (PROM) (Mokkink et al., 2018a). Although literature presents several aspects to evaluate the quality of measurement instruments, three domains should be highlighted: reliability, validity, and responsiveness. For each domain, more measurement properties exist (i.e., aspects related to quality of measurement instruments)(Prinsen et al., 2018a).

The evaluation of measurement properties when assessing a variable may reveal the strengths and weaknesses. The ability of measurement instruments to evaluate predicted outcomes involving clinical conditions (e.g., migraine) should be

based on knowledge about these tools to prevent bias. Thus, the correct characterization and identification of measurement instruments focused on the impacts of migraine is needed to know and apply measurement properties and direct new research about their impacts on patients with migraine (Terwee et al., 2007; Diamond et al., 2014; Mokkink et al., 2018a).

Therefore, this study aims to critically appraise, compare and summarize the quality of the measurement properties of all PROMs translated and validated for the Brazilian population that assess the impact of migraine.

2. Methodology

This systematic review was conducted according to a protocol previously registered in the International Prospective Register of Systematic Reviews (number CRD42021238930).

Eligibility criteria

We included studies performed with people diagnosed with migraine regardless of age and gender, without restriction of year of publication, which assessed the impact of migraine/quality of life as clinical outcome, investigated scales or questionnaires developed for clinical use in this population, evaluated at least one of the measurement properties (i.e., validity, reliability, and responsiveness), and were validated for Brazilian Portuguese.

Studies that use the PROM as an outcome measure were excluded, that is, studies in which the PROM is used to measure outcomes or used in a validation study of another instrument. Articles without full text were also excluded since it would hinder evaluating the quality of the study and results of measurement properties.

Search strategy

We conducted a literature search to identify studies evaluating measurement properties of PROMs developed to assess the impact of migraine and quality of life. Searches were conducted in MEDLINE/Pubmed, Web of Science, LILACS, and Embase databases. The search was carried out from February to June 2021. Table 1 shows the complete search strategy.

Table 1. Search strategy for each database to identify instruments adapted to the Brazilian population and designed to assess the impact of migraine.

MEDLINE/Pubmed	("Surveys and Questionnaires"[mh] OR Nonrespondent*[tiab] OR Questionnaire*[tiab] OR "Randomized Response Technique"[tiab] OR "Randomized Response Techniques"[tiab] OR Respondent*[tiab] OR "Response Technique"[tiab] OR "Response Techniques"[tiab] OR Survey*[tiab] OR "Techniques, Randomized Response"[tiab] OR "technique, delphi"[tiab] OR Psychometrics[mh] OR Psychometric*[tiab]) AND (Brazil[mh] OR Brasil*[tiab] OR Brazil*[tiab] OR portuguese[tiab]) AND ("Health Status Indicators"[mh] OR "functional status"[tiab] OR "Quality of Life"[mh] OR "HRQOL"[tiab] OR "Quality Of Life"[tiab] OR "Life Quality"[tiab] OR (Measurement[tiab] AND (propert*[tiab] OR valuation*[tiab])) OR ((Index*[tiab] OR Appraisal*[tiab] OR indicator*[tiab] OR Indice*[tiab]) AND Health[tiab])) AND ("Migraine Disorders"[mh] OR Migraine*[tiab] OR headache[tiab])
LILACS	(mh:"Surveys and Questionnaires" OR mh:psicometria OR tw:(Nonrespondent* OR Questionnaire* OR "Randomized Response Technique" OR "Randomized Response Techniques" OR Respondent* OR "Response Technique" OR "Response Techniques" OR Survey* OR "Techniques, Randomized Response" OR "technique, delphi" OR Psicometria OR Psychometric*)) AND (mh:Brazil OR tw:(Brasil* OR Brazil* OR portuguese)) AND (mh:"Health Status Indicators" OR tw:"functional status" OR mh:"Qualidade de Vida" OR tw:"Qualidade de Vida" OR tw:"HRQOL" OR tw:"QVRS" OR tw:"CVRS" OR tw:"Calidad de Vida" OR mh:"Quality of Life" OR tw:"HRQOL" OR tw:"Quality Of Life" OR tw:"Life Quality" OR (tw:Measurement AND (tw:propert* OR tw:valuation*)) OR ((tw:Index* OR tw:Appraisal* OR tw:indicator* OR tw:Indice*) AND tw:Health)) AND (mh:"Migraine Disorders" OR tw:Migraine* OR tw:headache)
	(mh:"Surveys and Questionnaires" OR mh:psicometria OR tw:(Nonrespondent* OR Questionnaire* OR "Randomized Response Technique" OR "Randomized Response Techniques" OR Respondent* OR "Response

SCIELO	Technique" OR "Response Techniques" OR Survey* OR "Techniques, Randomized Response" OR "technique, delphi" OR Psicometria OR Psychometric*)) AND (mh:Brazil OR tw:(Brasil* OR Brazil* OR portuguese)) AND (mh:"Health Status Indicators" OR tw:"functional status" OR mh:"Qualidade de Vida" OR tw:"Qualidade de Vida" OR tw:"HRQOL" OR tw:"QVRS" OR tw:"CVRS" OR tw:"Calidad de Vida" OR mh:"Quality of Life" OR tw:"HRQOL" OR tw:"Quality Of Life" OR tw:"Life Quality" OR (tw:Measurement AND (tw:propert* OR tw:valuation*)) OR ((tw:Index* OR tw:Appraisal* OR tw:indicator* OR tw:Indice*) AND tw:Health)) AND (mh:"Migraine Disorders" OR tw:Migraine* OR tw:headache)
EMBASE	('Surveys and Questionnaires':ti,ab OR Questionnaire*:ti,ab OR Nonrespondent*:ti,ab OR respondent*:ti,ab OR Questionnaire/exp OR 'Randomized Response Technique':ti,ab OR 'Randomized Response Techniques':ti,ab OR Respondent*:ti,ab OR 'Response Technique':ti,ab OR 'Response Techniques':ti,ab OR Survey*:ti,ab OR 'Techniques, Randomized Response':ti,ab OR 'technique, delphi':ti,ab OR Psychometric*:ti,ab) AND (Brazil/exp OR Brasil*:ti,ab OR Brazil*:ti,ab OR portuguese:ti,ab) AND ('Health Status Indicator'/exp OR 'Health Status Indicator':ti,ab OR 'Health Status Indicators':ti,ab OR 'functional status':ti,ab OR "Quality of Life":ti,ab OR "HRQOL":ti,ab OR "Life Quality":ti,ab OR (Measurement:ti,ab AND (propert*:ti,ab OR valuation*:ti,ab)) OR ((Index*:ti,ab OR Appraisal*:ti,ab OR indicator*:ti,ab OR Indice*:ti,ab) AND Health:ti,ab)) AND (Migraine/exp OR Migraine*:ti,ab OR headache:ti,ab)
Web of Science	TS=(("Surveys and Questionnaires" OR Nonrespondent* OR Questionnaire* OR "Randomized Response Technique" OR "Randomized Response Techniques" OR Respondent* OR "Response Technique" OR "Response Techniques" OR Survey* OR "Techniques, Randomized Response" OR "technique, Delphi" OR Psychometric*) AND (Brasil* OR Brazil* OR Portuguese) AND ("HRQOL" OR "Quality Of Life" OR "Life Quality" OR "Status Indicators" OR "functional status" OR (Measurement AND (property* OR valuation*)) OR ((Index* OR Appraisal* OR indicator* OR Indice*) AND Health)) AND (Migraine* OR headache))

Source: Authors.

Two independent reviewers (M.M. and H.F.) conducted the search, evaluated the information, and verified eligibility and selection criteria of studies found in the databases and gray literature.

Study selection

This systematic review followed recommendations for evaluating international studies using the delphi methodology of the COSMIN Checklist (Mokkink et al., 2019). Two reviewers (M.M. and H.F.) identified and selected studies by reading titles and abstracts, according to eligibility criteria. During the screening and selection process, duplicate studies were removed and a reference list was created for full text reading. The included reviews were read in full and the data were subsequently extracted. In any disagreement, a third researcher (RC) was present to resolve the disagreements.

Data extraction

After study selection, all relevant data were extracted to analyze measurement properties. Measurement instruments were evaluated from PROMs based on the COSMIN guideline developed to analyze the methodological quality of studies, such as systematic reviews of PROMs (Mokkink et al., 2018a). Data extracted were reliability (internal consistency, test-retest, and intra- and inter-rater), validity (content, construct, and known groups), responsiveness, interpretation (minimum important change or minimum important difference), and accuracy (data quality and final effects). Evidence for functional properties included acceptability (relevance and respondent burden) and feasibility (Mokkink et al., 2018a; Haywood et al., 2018).

Methodological quality assessment

Two independent reviewers (M.M. and H.F.) assessed the methodological quality of studies using the COSMIN checklist (Table 2). A third reviewer (R.C.) was consulted in case of disagreements.

COSMIN checklist

The COSMIN checklist was developed based on the Delphi methodology, which aims to create consensus(Diamond

et al., 2014) related to measurements properties, norms, and how the study design and statistical analysis should be evaluated (Mokkink et al., 2018a). The main objective of COSMIN is to assess the methodological quality of studies involving health-related patient-reported outcomes. We analyzed relevant results using the COSMIN checklist and collected transparent and systematic information from each measurement property.

Risk of bias

The COSMIN risk of bias checklist was used to assess the methodological quality and risk of bias of each article(Mokkink et al., 2019). As studies describe and evaluate different measurement properties, the COSMIN uses modular tools that must be completed for each measurement property of the study. According to COSMIN, the lowest score among items of a specific measurement property is considered for assessing the quality of the study, which can be classified as good, adequate, doubtful, inadequate, or not applicable. In this sense, if a specific measurement property is considered "inadequate", this specific item of the study has an "inadequate" general classification(Prinsen et al., 2018b).

The measurement properties considered in this study were structural validity (degree to which the score of a PROM is adequate), internal consistency (relevance of the interrelationship between determined items), cross-cultural validity/measurement invariance (performance of culturally adapted items), reliability (proportion of total variance attributed to true differences between patients), measurement error (systematic and random error of each patient), criterion validity (degree of PROMs score and its adaptations), hypothesis testing for construct validity (degree to which the score of a PROM is consistent with its hypotheses), and responsiveness (if the responsiveness of a PROM detects changes over time in the construct to be measured).

Results of measurement properties were selected and separated quantitatively and qualitatively. These results were also compared with criteria for good measurement properties (rating) to determine whether the measurement property was sufficient (+), insufficient (-), or indeterminate (?)(Mokkink et al., 2018a). The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) classified abstracts and quality of evidence.

Quality of evidence

Risk of bias (quality of studies), inconsistency of results, indirect evidence (evidence from different populations, interventions, or results of interest), and inaccuracy (total of samples included in the study) must be considered when analyzing measurement properties of systematic reviews of PROMs. The modified GRID approach (high, moderate, low, and very low) was applied to assess the quality of evidence (Mokkink et al., 2019).

Interpretation and feasibility

Interpretation and feasibility are the degree of qualitative significance (e.g., clinical connotations) to an individual PROM score or change in PROM scores. Sometimes, information about the distribution of scores is needed to interpret measurement properties, reveal groups of scores, and indicate floor and ceiling effects. The feasibility of an instrument is related to cost and application time and can be assessed in patients who complete the PROM and the professional who uses the PROM. Interpretation and feasibility are important aspects for the selection of instruments despite not being considered measurement properties (i.e., do not refer to the quality of PROMs). (Prinsen et al., 2018b).

Data analysis

Data were systematized and described in tables to summarize results obtained from each question. Therefore, we used tables of risk of bias and characterized the studies selected.

3. Results

The initial search found 112 articles, of which seven were selected for full reading, and four (three PROMs) were included in the review. (Figure 1).

Figure 1. COSMIN flowchart, 2018.
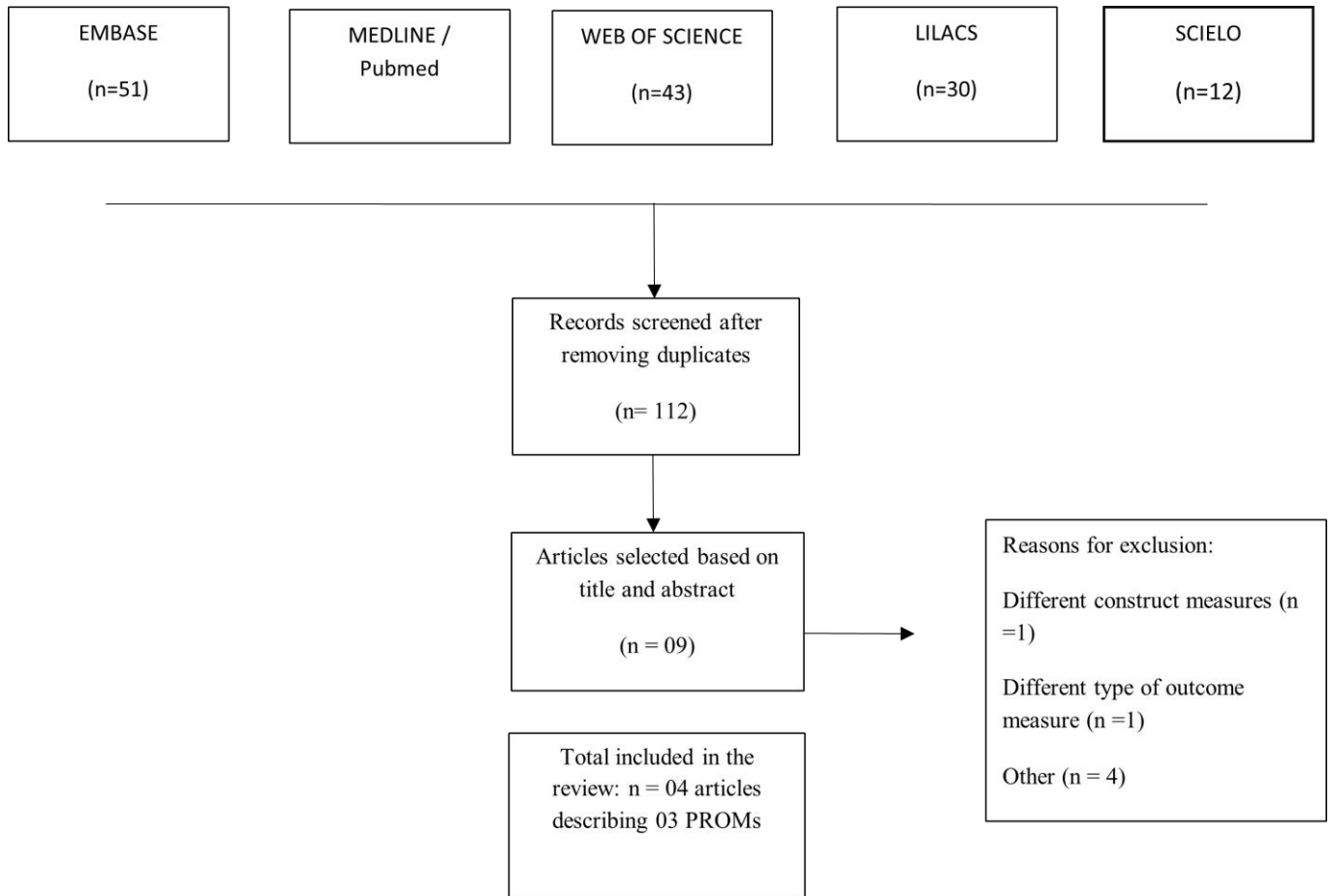
Source: COSMIN (2018).

The selected studies assessed three PROMs adapted to Brazilian Portuguese: Short-Form Headache Impact Test (HIT-6)(Martin et al., 2004a; Pradela et al., 2021b), Pediatric Migraine Disability Assessment (PedMIDAS)(Sampaio Rocha-Filho; Hershey, 2017), and Headache Disability Inventory (HDI)(Pradela et al., 2021a). Tables 2 and 3 show the characteristics of these PROMs and the population included in the studies. None of these instruments were developed in Brazil. (Tables 2 and 3)

Table 2. Characteristics of the three PROMs included. Methodological quality assessment using the COSMIN checklist of studies that assess the impact of migraine and adapted to the Brazilian population.

PROM* (a reference to the first article)	Construct(s) Target	Population	Mode of administration	Recall period	(Sub)scale (s) (number of items)	Response options	Range of scores Scoring	Original language	Available translations
Short-Form Headache Impact Test (HIT-6)(MARTIN et al., 2004a)	Internal consistence	1204 patients	Self-completed	Present time and four weeks.	06 items	Scores are computed by assigning a value of 6 to a response of "Never," 8 to "Rarely," 10 to "Sometimes," to "Very Often," and 13 to "Always." There are no "right" or "wrong" answers, so please put down your best guess.	A score of 49 or less reflected little or no impact; a score between 50 and 55 reflected some headache impact; and a score of 60 or more reflected severe headache impact the patient's ability to function in everyday life.	English	United States (English), Belgium (French, Flemish), Brazil (Portuguese), Canada (English, French), Finland (Finnish), France (French), Germany (German), Greece (Greek), Hungary (Hungarian), Israel (Hebrew), Mexico (Spanish), Netherlands (Dutch), Slovakia (Slovakian), South Africa (English), and United Kingdom (English).
Pediatric Migraine Disability Assessment (PedMIDAS)(SAM PAIO ROCHA-FILHO; HERSHEY, 2017)	Cross-cultural validity, test/retest reliability, Internal consistence.	100 patients	Self-report interview-based, parent/proxy report	Last three months	The are six questions		The PedMIDAS is scored by summing the answers across the six questions. Disability Grade 0 to 10 (Little to none) to Greater than 50 (Severe)	English	English and Brazil (Portuguese)

Headache Disability Inventory (IHD-Brazil)(PRADELA et al., 2021a)	Cross-cultural validity, reliability (internal consistence, intra-rater, and test)	30 patients	Self-report	Present time	The are 25 questions	Yes (4); sometimes (2) and No (0).	Its total score, ranging from 0–100 points, ranks the individual from absence to the maximum level of disability	English	Spanish, German, and Portuguese
Headache Impact Test (HIT-6™ Brazil)(PRADELA et al., 2021b)	Cross-cultural validity, reliability (internal consistence, intra-rater, and test)	132 patients	Self-report	last 30 days	06 items	For each question, there are four answer options. Never (6 points each), rarely (8 points each), sometimes (10 points each), very often (11 points each), and always (13 points each).	Score: < 60 Your headaches are having a very severe impact on your life. >49 Your headaches seem to be having little to no impact on your life at this time.	English	United States (English), Belgium (French, Flemish), Brazil (Portuguese), Canada (English, French), Finland (Finnish), France (French), Germany (German), Greece (Greek), Hungary (Hungarian), Israel (Hebrew), Mexico (Spanish), Netherlands (Dutch), Slovakia (Slovakian), South Africa (English), and United Kingdom (English).

Source: COSMIN (2018).

Table 3. Sample characteristics of the included studies.

PROMs	N	Age Mean (SD, range) yr	Gender n (%) Females	Disease	Disease duration	Disease severity	Setting	Country	Language
Short-Form Headache Impact Test (HIT-6)(MARTIN et al., 2004a)	1204	Participants were between 18 and 65 years of age (average 40 years), 84% were female, and 87% were caucasian.	84% were female, and 87% were caucasian.	Migraine	On average, subjects had a history of headaches per month.	About 81% had been previously diagnosed with migraine, and 55% reported their usual headache as severe.	A multicenter, international cross-sectional study conducted in a primary care setting. Data obtained from 1,171 adults from 14 countries.	Brazil	Portuguese
Pediatric Migraine Disability Assessment (PedMIDAS)(SAMPAIO ROCHA-FILHO; HERSHEY, 2017)	100	Psychometric assessment, the mean age was 11 6 2.1 years (range from 6 to 16)	47% females	Migraine	≥ 3 months	That the subjective impression of the disability exceeded both the frequency and severity influence	Child neurology clinic of the University Hospital Oswaldo Cruz.	Brazil	Portuguese
Headache Disability Inventory (IHD-Brazil)(PRADELA et al., 2021a)	30	Age 34.9, SD 11.5 years old	Total sample, 90% (n ¼ 27) were women.	Migraine	≥ 3 months	Acute or chronic	Tertiary headache outpatient clinic in Ribeirão Preto, Brazil	Brazil	Portuguese
Headache Impact Test (HIT-6™ Brazil)(PRADELA et al., 2021b)	132	Mean 44.3 (12.1 SD)	74.6% females	Migraine	≥ 3 months	chronic	Tertiary headache outpatient clinic in Ribeirão Preto, Brazil	Brazil	Portuguese

N=number of patients. Source: COSMIN (2018).

Table 4 shows the methodological quality assessment of studies regarding the criteria for good measurement properties. No study analyzed content validity and measurement error. Criterion validity was evaluated using construct validity because no gold standard is defined in the literature.

Table 4. Methodological quality assessment of studies analyzed for good measurement properties.

Instrument/ Author year)	Content validity			Structural validity			Internal consistency			Cross-cultural validity/measurement invariance			Reliability			Measurement error			Criterion validity			Hypothesis testing for construct validity			Responsiveness		
	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)	n	Meth Qual	Result (rating)
Short-Form Headache Impact Test (HIT-6)(MARTIN et al., 2004a)	-	-	-	-	-	-	44	Very good Cronbach's alpha 0.82	(+)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pediatric Migraine Disability Assessment (PedMIDAS) (SAMPAIO ROCHA-FILHO; HERSHEY, 2017)							100	Very good Cronbach's alpha 0.84	(+)	40	Inadequate	(?)	24	Inadequate (No Kappa or ICC calculated)	(?)				-	-	-	-	-	-
Headache Disability Inventory (IHD-Brazil)(PRADELA et al., 2021a)	-	-	-	132	Adequated (least 5 the times number of items, and ≥100)	? (IRT/Rasch: Model fit not reported)	30	Very good Cronbach's alpha 0.84	(+) All domains above 0.70	30	Inadequate <100 subjects per group	(?)	67	very good ICC=0.95	(+)	-	-	-	-	-	-	132	Very good	(-) Correlation with gold standard < 0.70 OR AUC <0.70 (r=0.67)	-	-	-
Headache Impact Test (HIT-6™ Brazil)(PRADELA et al., 2021b)	-	-	-	132	Adequated (least 5 the times number of items, and ≥100)	? (IRT/Rasch: Model fit not reported)	132	Very good Cronbach's alpha 0.97	(+)	-	-	-	67	very good ICC=0.95	(+)	-	-	-	-	-	-	132	Very good	(-) Correlation with gold standard < 0.70 OR AUC <0.70 (r=0.64)	-	-	-

"+" = sufficient, " –" = insufficient, "?" = indeterminate, ICC = intraclass correlation coefficient, ES = Effect Size, IC = Confidence interval, ROC =receiver operating characteristic, MIC = minimal important change, LoA = limits of agreement, SEM = Standard Error of Measurement, SDC = smallest detectable change. Source: COSMIN (2018).

Internal consistency

Internal consistency was assessed by the three PROMs included in this study. Methodological quality was classified as very good, and Cronbach's alpha ranged from 0.82 to 0.97.

Cross-cultural validity/measurement invariance

Only two PROMs (PedMIDAS and HDI) performed cross-cultural validity. They were classified as inappropriate since regression or confirmatory factor analyses were not used, and samples were lower than 100 individuals.

Reliability

Reliability was classified as very good in two PROMs (HIT-6 and HDI) (intraclass correlation coefficient [ICC] of 0.95). In one study (PedMIDAS), the PROM was classified as inadequate for not reporting ICC or Kappa.

Hypothesis testing for construct validity

The hypothesis testing for construct validity was classified as very good in two studies (HDI and HIT-6) since they clearly informed the comparator instrument and performed appropriate analyses.

Responsiveness

No study evaluated this measurement property.

Floor and ceiling effects

Although floor and ceiling effects may result in insufficient reliability and feasibility, they are descriptive and not considered a formal measurement property(Prinsen et al., 2018b). Only one study evaluated floor and ceiling effects (HIT-6). The ceiling effect was achieved in the HIT-6 questionnaire (9.1% of respondents).

Quality of evidence

Quality of evidence of the evaluated PROMs was summarized and classified in Table 5.

Table 5. Quality of evidence of the three PROMs assessed, according to the modified GRADE approach.

PROM	Studies (n)	Ref	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality of evidence
Headache Impact Test (HIT-6™ Brazil)(PRADELA et al., 2021b)˙(MARTIN et al., 2004a)	2	MARTINI et al, 2004; PRADELA et al, 2020	-1 serious [1]	Moderate (not downgrade)	-	-	Moderate
Pediatric Migraine Disability Assessment (PedMIDAS)(SAMPAIO ROCHA-FILHO; HERSHEY, 2017)	1	SAMPAIO et al, 2017	-3 serious [2]	-	-	-	Very Low
Headache Disability Inventory (IHD-Brazil)(PRADELA et al., 2021a)	1	PRADELA et al, 2020	-2 serious [3]	-	-	-	Low

(n) = number of study
Note: Criterion validity was considered for construct validity.
(1) we lowered 1 level, as the questionnaires did not evaluate (Cross-cultural validity/measurement invariance, content validity, Criterion validity, Measurement error and responsiveness).
(2) we lowered three levels, as the questionnaire did not evaluate (content validity, Criterion validity, Measurement error, Hypothesis testing for construct validity, and responsiveness). Cross-cultural validity and reliability were "inadequate".
(3) we lowered two levels, as the questionnaire did not evaluate (content validity, Criterion validity, Measurement error and responsiveness). Cross-cultural validity was "inadequate. Source: COSMIN (2018).

4. Discussion

This review systematically gathers evidence from four studies and critically evaluates measurement properties of PROMs using the COSMIN (Martin et al., 2004b; Sampaio Rocha-Filho; Hershey, 2017; Pradela et al., 2021a; Pradela et al., 2021b;). All instruments found were developed to assess the impact of migraine and were validated for Brazil. Instruments presented good internal consistency and adequate structural validity; PedMIDAS did not evaluate the latter measurement property. HDI and PedMIDAS presented inadequate cross-cultural validity. PedMIDAS presented inadequate reliability. Content validity, measurement error, criterion validity, and responsiveness were not analyzed in any of included studies.

The only measurement property evaluated in all included studies was internal consistency, classified as very good. The unidimensionality of scales or subscales is a prerequisite for analyzing internal consistency in questionnaires based on reflective models; the analyzed items presented these characteristics and a high Cronbach's alpha. Regarding clarity of items, the latest version of COSMIN suggests that a factorial or theoretical analysis of responses should be performed. Only the study using PedMIDAS did not perform factor analysis(Prinsen et al., 2018b; Mokkink et al., 2018b).

The analysis of cross-cultural validity/measurement invariance is needed to verify the degree of performance of items of a cross-culturally adapted instrument and how much the translated instrument is an adequate reflection of the original version. This measurement property was classified in PedMIDAS and HDI as inadequate methodological quality since COSMIN advises that multiple group factor or differential item functioning analyses should be performed(Prinsen et al., 2018b). These analyses aim to measure if different groups respond similarly to a specific item.

The sample evaluated in these studies was homogeneous, with the population having important characteristics and other similarities, such as diagnosis, disease severity, educational level, and age group. However, even with these similarities, the COSMIN indicates the need for statistical methods based on the classical test theory or item response theory/Rasch

measurement (Prinsen et al., 2018b). Although large samples would be more appropriate for increasing the reliability of confirmatory factor analysis, the studies that evaluated this measurement property followed recommendations to use samples of 30 to 50 individuals(Prinsen et al., 2018b).

Regarding risk of bias, the three PROMs were applied to a homogeneous group of stable patients under correct conditions. The authors also considered an appropriate interval of two weeks to avoid memory bias and ensure that conditions of patients did not change the measured construct.

The statistical analysis that assess reliability is ICC, in which the proportion of total variance in the measurement attributable to true differences between patients is considered. On the other hand, the weighted Kappa statistical method is appropriate for continuous and ordinal scores(Mokkink et al., 2018b). HIT-6 and HDI were classified as adequate since both presented ICC = 0.95. PedMIDAS was classified as inadequate since none of the methods for good measurement property were used.

The measurement error refers to systematic and random errors of each patient, and this score is not attributed to real changes in the measured construct. The standard error of measurement (SEM) is the most appropriate statistic for anayzing the measurement erros of studies based on the classical test theory and using test-retest. The calculation of SEM from Cronbach's alpha is inappropriate because it does not consider the variation between moments. Other widely used statistical methods are the limits of agreement and the smallest detectable change. SEM was not found in the three PROMs included in this review(Prinsen et al., 2018b). Although Kappa statistic is a measure of reliability, authors commonly consider it a measure of agreement(Mokkink et al., 2018a). From a clinical point of view, the absence of this measure requires strong and adequate evidence of reliability for justifying its relevance(Winser et al., 2019).

Responsiveness is the capacity of a PROM to detect changes over time in the model to be measured; therefore, it refers to the validity of a score change. Although this analysis was not performed in any of the PROMs included in this review, HDI and HIT-6 performed a construct/hypothesis validity test in which the only difference from responsiveness was that validity refers to a single score. Even though the gold standard was not defined in the two studies analyzed, they tested hypotheses to analyze the expected magnitude of an intervention. Therefore, according to COSMIN, HDI and HIT-6 were classified as very good(Mokkink et al., 2018a; Pradela et al., 2021b; Pradela et al., 2021a).

Four factors were considered for analyzing quality of evidence: risk of bias, inconsistency, inaccuracy, and indirectness. HIT-6 presented a moderate quality of evidence, with serious risk of bias and moderate inconsistency. PedMIDAS presented a very low quality of evidence, with extremely serious risk of bias. Last, HDI presented low quality of evidence with very serious risk of bias.

Therefore our results reaffirm the importance of studies evaluating measurement properties to increase methodological quality(Terwee et al., 2012) but do not discard the use of these instruments. Nevertheless, we recommend using the COSMIN to verify instruments in detail and increase reliability in scientific research and clinical practice.

The Migraine Disability Assessment was not included in this study despite being a relevant instrument widely used in the Brazilian population(Ferreira et al., 2021). According to our search strategy, this questionnaire did not meet the eligibility criteria because no study analyzed its measurement properties or validated for Brazilian Portuguese (Stewart et al., 2000; Ferreira et al., 2021).

5. Conclusion

According to the COSMIN, the three PROMs analyzed in this study presented important limitations. These limitations hampered a good quality of evidence of instruments used to assess the impact of migraine. Nevertheless, HIT-6 was the most recommended because it presented a moderate quality of evidence.

The latest update to the COSMIN guidelines was published following the publication of two of the four included studies; the analysis that can influence the verification quality results. Thus, it is suggested that new studies be conducted based on the new COSMIN guideline, aiming at better measures of instrument results and increasing the reliability of selected tools for research and clinical practice.

References

Agosti, R. (2018) Migraine Burden of Disease: From the Patient's Experience to a Socio-Economic View. *Headache*, 58: 17–32. https://doi.org/10.1111/head.13301.

Diamond, I. R. et al. (2014) Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*. 67 (4): 401–409. doi: 10.1016/j.jclinepi.2013.12.002.

Edwards, R. R. et al. (2006) Catastrophizing and pain in arthritis, fibromyalgia, and other rheumatic diseases. *Arthritis Care and Research*. 55 (2): 325–332. doi: 10.1002/art.21865.

Ferreira, P. L. et al. (2021) Acesso livre Validade e confiabilidade da versão em português do Migraine Disability Assessment modificado. *BMC neurology* 1–11. doi: 10.1186/s12883-021-02085-z.

Goadsby, P. J. et al. (2017) Pathophysiology of Migraine: A Disorder of Sensory Processing. *Physiological Reviews*. 97 (2): 553–622, abr. doi: 10.1152/physrev.00034.2015.

Guyatt, G. H. et al. (2011) GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology*, 64 (4): 380–382. doi: 10.1016/j.jclinepi.2010.09.011.

Haywood, K. L. et al. (2018) Assessing the impact of headaches and the outcomes of treatment: A systematic review of patient-reported outcome measures (PROMs). *Cephalalgia*. 38 (7):1374–1386. doi: 10.1177/0333102417731348.

Martin, M. et al. (2004a) The Short-Form Headache Impact Test (HIT-6) was psychometrically equivalent in nine languages. *Journal of Clinical Epidemiology*. 57 (12) :1271–1278. doi: 10.1016/j.jclinepi.2004.05.004.

Martin, M. et al. (2004b) The Short-Form Headache Impact Test (HIT-6) was psychometrically equivalent in nine languages. *Journal of Clinical Epidemiology*. 57 (12): 1271–1278, dez. doi: 10.1016/j.jclinepi.2004.05.004.

Mokkink, L. B. et al.(2018a) COSMIN manual for systematic reviews of PROMs, user manual. n. February. 1–78. Retrieved from https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf.

Mokkink, L. B. et al. (2018b) COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*. 27(5):1171-1179. doi: 10.1007/s11136-017-1765-4.

Mokkink, L. B. et al. (2019) COSMIN Study Design checklist for Patient-reported outcome measurement instruments. n. July. Retrieved from https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf.

Pradela, J. et al. (2021a) Brazilian Portuguese version of the Headache Disability Inventory: Cross-cultural adaptation, validity, and reliability. *Cephalalgia*. 41 (2):156–165. doi: 10.1177/0333102420949860.

Pradela, J. et al. (2021b) Measurement properties of the Headache Impact Test (HIT-6[TM] Brazil) in primary and secondary headaches. Headache. 61 (3): 527–535. doi: 10.1111/head.14049.

Prinsen, C. A. C. et al. (2018a) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. (27) 5: 1147–1157. doi: 10.1007/s11136-018-1798-3.

Prinsen, C. A. C. et al. (2018b) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 27 (5): 1147–1157, 12 maio. doi: 10.1007/s11136-018-1798-3

Queiroz, L. P.; Silva Junior, A. A. (2015) The prevalence and impact of headache in Brazil. *Headache*. 55 (S1): 32–38. doi: 10.1111/head.12511.

Sampaio Rocha-Filho, P. A.; Hershey, A. D. (2017) Pediatric Migraine Disability Assessment (PedMIDAS): Translation Into Brazilian Portuguese and Cross-Cultural Adaptation. *Headache*. 57 (9): 1409–1415. https://doi.org/10.1111/head.13159.

Steiner, T. J. et al. (2018) Migraine is first cause of disability in under 50s : will health politicians now take notice ? *J Headache Pain* Feb 21;19(1):17.

 doi: 10.1186/s10194-018-0846-2..

Steiner, T. J. et al.(2020) Migraine remains second among the world's causes of disability, and first among young women: findings from GBD2019. The *Journal of Headache and Pain*. 21 (1): 137, 2 dez. doi: 10.1186/s10194-020-01208-0.

Stewart, W. F. et al. (2000) Validity of the Migraine Disability Assessment (MIDAS) score in comparison to a diary-based measure in a population sample of migraine sufferers. *Pain*. 88 (1): 41–52. doi: 10.1016/S0304-3959(00)00305-5.

Terwee, C. B. et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 60 (1): 34–42. doi: 10.1016/j.jclinepi.2006.03.012.

Terwee, C. B. et al. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*. 21 (4): 651–657, 6 maio. doi: 10.1007/s11136-011-9960-1.

Winser, S. J. et al. (2019) Measures of balance and falls risk prediction in people with Parkinson's disease: a systematic review of psychometric properties. *Clinical Rehabilitation*. 33 (12):1949–1962, 1 dez. doi: 10.1177/0269215519877498.

Yang, M. et al. (2011) Validation of the Headache Impact Test (HIT-6™) across episodic and chronic migraine. *Cephalalgia*. 31 (3): 357–367, 6 fev. doi: 10.1177/0333102410379890.