
Article:
Venda Nova, C., Zakrzewska, J.M., R Baker, S. et al. (1 more author) (2021) Patient reported outcome measures in trigeminal neuralgia - a systematic review of psychometric performance. European Journal of Pain. ISSN 1090-3801

https://doi.org/10.1002/ejp.1779

Received: 11 November 2020 | Revised: 7 April 2021 | Accepted: 8 April 2021

DOI: 10.1002/ejp.1779

REVIEW ARTICLE

EJP European Journal of Pain

Patient reported outcome measures in trigeminal neuralgia – A systematic review of psychometric performance

Carolina Venda Nova[1] | **Joanna M. Zakrzewska**[1] | **Sarah R. Baker**[2] | **Richeal Ni Riordain**[1,3]

[1]Eastman Dental Institute, University College London, London, UK

[2]School of Clinical Dentistry, Sheffield University, Sheffield, UK

[3]Cork University Dental School and Hospital, Cork, Ireland

Correspondence
Carolina Venda Nova, Eastman Dental Institute, University College London, Rockefeller Building, 21 University Street, London WC1E 6DE, UK.
Email: carolina.nova.18@ucl.ac.uk

Funding information
This work was supported by the Trigeminal Neuralgia Association UK (TNAUK) and by the Rosetrees Trust (grant A2327). J.Z. undertook this work at UCL/UCLHT, who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centre funding scheme. The founding sources had no role in the study design, in the collection, analysis and interpretation of the data, in the writing of the report and in the decision to submit the article for publication.

Abstract

Background: Trigeminal neuralgia (TN) is a rare condition for which there are multiple treatment options available. To date, there has been difficulty in comparing the outcomes of treatment due to the variety of patient-reported outcome measures (PROMs) and their inadequate psychometric testing. The aim of this review was to assess the psychometric properties of PROMs used to date in TN and make recommendations for their use in future studies.

Methods: Five electronic databases (MEDLINE, EMBASE, CINAHL Plus, PsycINFO, Health and Psychosocial Instruments) were searched for studies assessing the development of PROMs or their psychometric properties in TN studies. The records obtained were assessed independently by two reviewers for their methodological quality, following guidance from the Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN).

Results: Six studies were identified and information on five PROMs (Brief Pain Inventory Facial (BPI-F), Visual Analogue Scale (VAS), Barrow Neurology Institute Pain Scale (BNI-PS), Penn Facial Pain Scale-Revised (Penn-FPS-R) and Trigeminal Neuralgia Quality of Life Score) were retrieved. The Penn-FPS-R demonstrated moderate quality evidence for sufficient content validity. The BPI-F showed moderate evidence for sufficient internal consistency and structural validity but low evidence for inconsistent content validity. The Trigeminal Neuralgia Quality of Life score showed very low-quality evidence for insufficient content validity, structure validity and responsiveness. No evidence was found on the assessment of any psychometric properties of the VAS and BNI-PS in TN.

Conclusion: There is limited evidence of the psychometric performance of patient-reported outcomes for TN and recommendations for their inclusion in future studies cannot be made. The validation of PROMs in TN studies should be a priority in this field of research.

Significance: This review highlights the knowledge gap in the field of psychometrics of patient reported outcomes measures in the field of TN. Given the unavailability of an objective outcome measure for pain or health related quality of life,

psychometrically sound PROMs are essential for assessing medical and surgical treatment outcomes in TN.

1 | INTRODUCTION

Trigeminal neuralgia (TN) is defined by the International Classification of Headache disorders (Olesen, 2018) and by the International Classification of Orofacial Pain (ICOP, 2020) as 'A disorder characterized by recurrent unilateral brief electric shock-like pain, abrupt in onset and termination, limited to the distribution of one or more divisions of the trigeminal nerve and triggered by innocuous stimuli'.

Despite being a rare condition, there are multiple surgical and pharmacological options available for its management and most of the outcomes of treatment collected in different studies have used measures that have not been psychometrically tested (Zakrzewska & Relton, 2016).

Over the past two decades, the biopsychosocial model of chronic pain has been widely promoted for the assessment and management of chronic pain (Gatchel et al., 2007). This has alerted the field to move beyond the exclusive assessment of pain intensity to allow incorporation of other domains which might be more meaningful to patients, such as the impact of pain on quality of life (QOL) (Sullivan & Ballantyne, 2016). Similar recommendations were made in the TN field (Cruccu et al., 2008) but there are scarce examples where it has happened (Cheng et al., 2017; Kotecha et al., 2017; Mousavi et al., 2016).

Given the subjectivity of constructs like pain and QOL, the direct reporting from the patient is of utmost importance. Patient-reported outcome measures (PROMs) are questionnaires or forms completed by the patients about their health without interpretation by a clinician or researcher (Weldring & Smith, 2013). A patient-reported outcome (PRO) can also be a record obtained by direct questioning or interviewing of the patient.

Patient-reported outcome measures should be chosen on their psychometric performance in the studied population to allow for comparison of study results. There have been repeated calls that measures of PRO assessment should be standardized and validated, exemplified by the Big Data for Better Outcomes comprehensive European research programme (IMI, 2018).

As a minimum, a questionnaire should be validated to be used in the target population for the outcome of interest, on a specific context, and should therefore be relevant, comprehensive and comprehensible – content validity (Terwee et al., 2018). The instruments should also demonstrate adequate structural validity (the instrument scores reflect the construct to be measured), reliability (the scores of the instrument do not change when patients are stable, despite possible changes in the timing of the measurement and the instrument rater) and responsiveness (if there is a change in the construct of interest, for example, due to a new treatment, the instrument is able to detect it). Additionally, the questionnaires should also be easy to interpret and be feasible to use without causing excessive burden on patients or clinicians (Prinsen et al., 2018).

Due to the current lack of guidance on the most appropriate instruments to be used in TN studies, the aim of this systematic review was to summarize and evaluate the psychometric properties of outcome measures that have been developed or adapted for TN patients undergoing treatment and to make recommendations for their use in future studies.

2 | METHODS

The methodology adopted for this systematic review follows guidance from COSMIN (Prinsen et al., 2018) and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA statement) (Moher et al., 2009).

A protocol was prepared and registered in PROSPERO (CRD42020185914, 1 July 2020) before starting the systematic review process.

2.1 | Literature search

A systematic search was performed in MEDLINE (PubMed) (1966–2020), EMBASE (Ovid) (1980–2020), CINAHL Plus with Full Text (1937–2020), PsycINFO (Ovid) and Health and Psychosocial Instruments (1985–2020).

The search was designed to identify all studies where there was (a) development, evaluation and/or validation of measurement properties of (b) PROMs in (c) adult TN patient cohorts. A published and validated search filter designed for Medline was used, with a high sensitivity for retrieving results on measurement properties studies (Terwee et al., 2009). This filter was then adapted for search in EMBASE, CINAHL and PsycINFO. The search on the Health and Psychosocial Instruments database was limited to the target population, that is, TN.

The full search strategy can be found in Appendix S1A.

2.2 | Eligibility criteria

Studies were included with a TN patient cohort >18 years of age, which aimed to evaluate and/or validate measurement

properties of PROM(s), develop a PRO or evaluate the interpretability of a PRO. Only full-text articles reported in English were included. Studies that described the use of clinician reported outcomes only were excluded. In addition, conference abstracts, editorials and conference proceedings were also excluded. A choice was made not to search for any specific PROM or specify domains or dimensions of the PROM as it was anticipated that the search would not yield many results.

2.3 | Study records

The records identified were transferred to EndNote X9.2 (Clarivate Analytics) and duplicates were removed. CVN and JMZ independently screened the records by title and abstract. SRB, RNR and CVN independently screened the records based on full text. Disagreements were resolved with discussion. Once records were identified as eligible to be included, data were extracted (see below).

2.4 | Measurement properties

2.4.1 | Data extraction

Data were extracted by one author (CVN) using a preselected form based on those recommended by COSMIN (Mokkink et al., 2018) on both study details (study design, sample size, gender, age, TN classification and type of treatment) and PROM description (PROM, construct, mode of administration and psychometric properties under study). A second author (RNR) confirmed the data extracted for accuracy. There were no disagreements.

The measurement properties under study were labelled according to the guidance provided by COSMIN (Prinsen et al., 2018). This included evidence of the assessment of the following measurement properties: content validity, internal structure (structural validity and internal consistency) and the remaining properties (test–retest reliability, measurement error, criterion validity, construct validity and responsiveness). Content validity, which is defined by COSMIN as the degree to which the content of a PROM is an adequate reflection of the construct to be measured, is considered the most important measurement property.

The assessment of content validity was based on guidance from a recent Delphi study (Terwee et al., 2018), which recommends that well-designed PROM development studies should be taken into consideration in the assessment of content validity. Development studies, which use qualitative research methods, allow for direct patient input in different stages, such as concept elicitation, item generation, comprehensibility and comprehensiveness (Magasi et al., 2012).

2.4.2 | Assessment of the methodological quality of the studies

Two reviewers (CVN and RNR) independently assessed the included studies to evaluate their methodological quality and consensus was reached during an online meeting. The methodological assessment was done in three steps as recommended by COSMIN.

In step 1, the methodological quality was assessed using the risk of bias checklist (Prinsen et al., 2018). This checklist consists of a table which outlines all the measurement properties as well as the PROM development study characteristics against quality standards. There are four possible scores for each standard: 'very good', 'adequate', 'doubtful' or 'inadequate'. The overall score for the methodological quality of the study was taken by using the 'the worst score counts' principle. Details of this can be found in Appendix S1B.

In step 2, criteria for good measurement properties were applied by using the following quality ratings: 'sufficient' (+), 'insufficient' (−) or 'indeterminate' (?) (see Table 1). At this stage, the results of different studies, if consistent, are pooled together for assessment of the overall quality rating of each PROM as 'sufficient' (+), 'insufficient' (−), 'inconsistent' (±) or 'indeterminate' (?).

Finally, in step 3, the Grading of Recommendations Assessment, Development and Evaluation modified method was used to grade the overall quality of the evidence collected of each measurement property as 'high', 'moderate', 'low' or 'very low' (Prinsen et al., 2018; Terwee et al., 2018).

3 | RESULTS

The search identified 549 titles. After 141 duplicates were removed, 408 abstracts were screened. Of these, 18 full-text articles were screened but only six were included for the final analysis. Figure 1 displays the flowchart of the study records, with documented reasons for exclusion in different phases of the screening process.

A total of five PROMs were identified in the six articles: BPI-F – Brief Pain Inventory Facial (Lee et al., 2010; Sandhu et al., 2015), VAS – Visual Analogue Scale (Reddy et al., 2013, 2014), BNI-PS – Barrow Neurology Institute Pain Scale (Reddy et al., 2013, 2014), Penn-FPS-R – Penn Facial Pain Scale-Revised (Symonds et al., 2018) and the Trigeminal Neuralgia QOL Score (TN QOLS; Luo et al., 2019). See Table 2 for the characteristics of the included studies.

3.1 | Brief Pain Inventory Facial

The BPI-Facial (Lee et al., 2010) was intended to be designed as subscale adaptation of the Brief Pain Inventory

TABLE 1 Criteria for good measurement properties

Measurement property	Rating	Criteria
Structural validity	+	CTT CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 IRT/Rasch No violation of *unidimensionality*: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 *AND* No violation of *local independence*: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's <0.37 *AND* No violation of *monotonicity*: adequate looking graphs OR item scalability >0.30 *AND* Adequate *model fit* IRT: $\chi^2 > 0.001$ Rasch: infit and outfit mean squares ≥0.5 and ≤1.5 OR *Z*-standardized values >−2 and <2
	?	CTT: not all information for '+' reported IRT/Rasch: model fit not reported
	−	Criteria for '+' not met
Internal consistency	+	At least low evidence for sufficient structural validity AND Cronbach's α(s) ≥0.70 for each unidimensional scale or subscale
	?	Criteria for "At least low evidence for sufficient structural validity" not met
	−	At least low evidence for sufficient structural validity AND Cronbach's α(s) <0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted κ ≥0.70
	?	ICC or weighted Kappa not reported
	−	ICC or weighted κ <0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	−	SDC or LoA >MIC
Criterion validity	+	Correlation with gold standard ≥0.70 OR AUC ≥0.70
	?	Not all information for '+' reported
	−	Correlation with gold standard <0.70 OR AUC <0.70
Responsiveness	+	The result is in accordance with the hypothesis OR AUC ≥0.70
	?	No hypothesis defined (by the review team)
	−	The result is not in accordance with the hypothesis OR AUC <0.70

Note: "+", positive rating; "−", negative rating; "?", indeterminate rating.

Adapted from: COSMIN guideline for systematic reviews of patient-reported outcome measures (Prinsen et al., 2018).

Abbreviations: AUC, area under the curve; CFA, confirmatory factor analysis; CFI, comparative fit index; CTT, classical test theory; ICC, intraclass correlation coefficient; IRT, item response theory; LoA, limits of agreement; MIC, minimal important change; RMSEA, root mean square error of approximation; SDC, small detectable change; SRMR, standardized root mean residuals; TLI, Tucker–Lewis index.

(Cleeland & Ryan, 1994) to allow for the inclusion of seven extra questions specific to interference of pain related to the face. It included, for example, questions about interference of pain on eating a meal or on smiling, laughing or talking. The BPI was originally developed to be used in cancer pain as a pain intensity (severity) and pain interference tool (Cleeland & Ryan, 1994). Since its development, it has been widely used in different pain conditions, translated into different languages and validated to be used in different clinical and research contexts. The pain intensity subscale consists of four items rated on an 11-point scale (0–10) with anchors 'no pain' and 'pain as bad as you can imagine'. The pain interference subscale consists of seven items rated on an 11-point scale (0–10) with anchors 'does not interfere' and 'completely interferes'. The BPI-F subscale on interference (face) consisted of seven

FIGURE 1 Flowchart of search strategy results

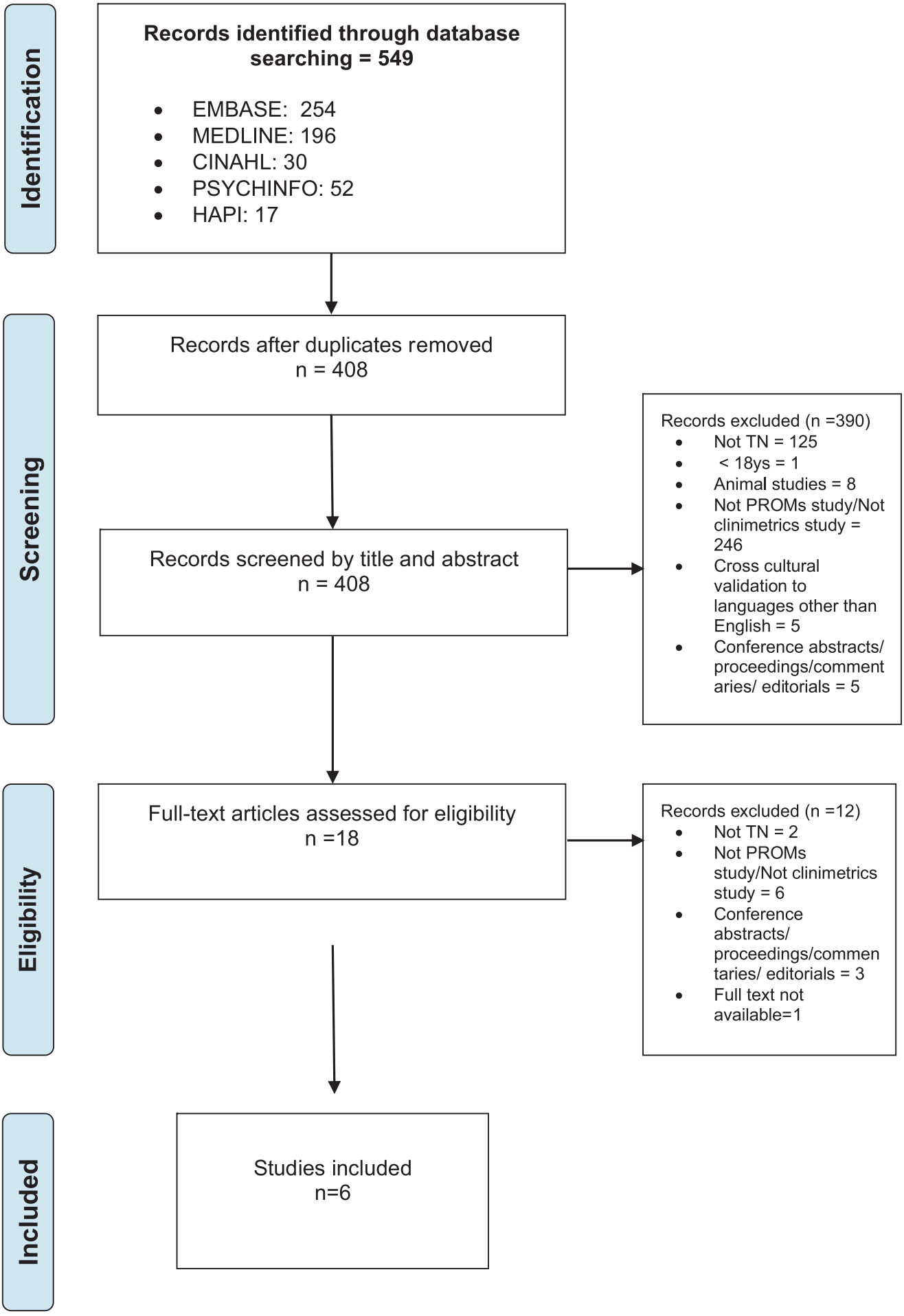
Identification
Records identified through database searching = 549
EMBASE: 254
MEDLINE: 196
CINAHL: 30
PSYCHINFO: 52
HAPI: 17
Screening
Records after duplicates removed n = 408
Records screened by title and abstract n = 408
Records excluded (n =390)
Not TN = 125
< 18ys = 1
Animal studies = 8
Not PROMs study/Not clinimetrics study = 246
Cross cultural validation to languages other than English = 5
Conference abstracts/ proceedings/comment aries/ editorials = 5
Eligibility
Full-text articles assessed for eligibility n =18
Records excluded (n =12)
Not TN = 2
Not PROMs study/Not clinimetrics study = 6
Conference abstracts/ proceedings/commen taries/ editorials = 3
Full text not available=1
Included
Studies included n=6

TABLE 2 Characteristics of included studies

Reference	Study design	TN classification	Sample size (%females)	Age, years $M \pm SD$	Treatment	PROM	Construct	Mode of administration	Psychometric properties evaluated[a]
Lee et al. (2010)	Cross sectional	Based on the International Headache Society Classification (2nd edition) and Burchiel Type 1 classic TN or Burchiel Type 2	156 (63%)	Type 1 – 61 (*SD* not available) Type 2 – 56 (*SD* not available)	Unclear	BPI – F	Pain interference facial	Self-completed by patient	Subscale development[a] Content validity Structural validity Internal consistency
Reddy et al. (2013)	Prospective cohort study	Based on the International Headache Society Classification (2nd edition)	60 (78%)	53.4 ± 12.3	MVD	VAS BNI-PS	Pain intensity	Face to face interviews at base line and at 2 years follow up	Interpretability[a]
Reddy et al. (2014)	Prospective cohort study	Based on the International Headache Society Classification (2nd edition)	43 (67%)	68.9 ± 11.5	Percutaneous stereotactic radiofrequency	VAS BNI-PS	Pain intensity	Face to face interviews at base line and 2 years follow up	Interpretability[a]
Sandhu et al. (2015)	Retrospective cohort study	Based on the International Headache Society Classification (2nd edition) Burchiel Type 1 classic TN or Burchiel Type 2	234 (62%)	62.1 ± 14.3	Neurosurgery	BPI – F	Pain intensity Pain interference general Pain interference facial	Self-completed by the patient at initial visit and 30 days after treatment	Interpretability[a]
Symonds et al. (2016)	Semi-structured interviews	Unclear	20 (85%)	57.5 ± 8.64	Medical treatment	Penn Facial Pain Scale Revised	Pain interference on HRQOL and daily functioning	Self-completed by the patient	Subscale development[a] Content validity
Luo et al. (2019)	Not described	Primary TN	298 (not available)	Not available	Radiofrequency thermocoagulation	TN QOLS	Quality of life	Self-completed by the patient	Subscale development[a] Content validity Criterion validity Structural validity Internal consistency Responsiveness

Abbreviations: BNI-PS, Barrow Neurological Institute Pain Scale; BPI-F, Brief Pain Inventory Facial; MVD, microvascular decompression; PROM, patient reported outcome measure; *SD*, standard deviation; TN QOLS, Trigeminal Neuralgia Quality of Life Score; VAS, Visual Analogue Scale.

[a]Subscale development and interpretability are not considered measurement properties but the former can be used to aid in content validity assessment and the latter should be assessed when the other measurement properties fulfil criteria of quality.

new items rated on an 11-point scale (0–10) with anchors 'does not interfere' and 'completely interferes'.

One study presented data on the BPI-F subscale development, structural validity and internal consistency (Lee et al., 2010), and one study presented data on the scale's interpretability (Sandhu et al., 2015).

3.1.1 | Validity

Subscale development and content validity

The subscale development study (Lee et al., 2010) was of doubtful quality as it is unclear if patients were asked about the comprehensibility and comprehensiveness of the PROM.

In the absence of content validity studies for this subscale, the content validity rating was based on the development study and on the reviewer's ratings which provided low evidence for inconsistent findings.

3.1.2 | Internal structure

Structural validity and internal consistency

One study of adequate quality assessed the structural validity and internal consistency of the BPI-Facial (Lee et al., 2010); however, it was not clear if the items of the subscale were based on a reflective or formative model but it was assumed that the items of the 'pain interference facial' construct were based on a reflective model drawn from the literature and consultation of experts in the field. The authors of the study hypothesized that the BPI-Facial could be a two- or three-factor questionnaire and conducted a principal factor analysis with varimax rotated factor. Three factors with eigenvalues >1 (interference facial 5.4/interference general 4.3/pain intensity 2.3) were identified and confirmed with a scree plot. The three factors explained 97.6% of the variance of the instrument. A cut-off >0.4 was used for the loading values suggesting a high correlation of the items with the domain. The pain interference facial factor loading varied from 0.73 (impact of pain on eating) to 0.87 (impact of pain on brushing and on smiling). These findings suggest moderate evidence for sufficient unidimensionality of the pain interference subscale.

The internal consistency of the pain interference facial subscale was 0.95 calculated using Cronbach's α. Taking into consideration the moderate evidence for sufficient structural validity and Cronbach's α >0.70, there is moderate evidence for sufficient internal consistency.

Interpretability

According to COSMIN, interpretability is not a measurement property, rather a feature to be taken into consideration when choosing an instrument as it attributes meaning to an instrument's single score or change in scores (Prinsen et al., 2018). One study assessed the interpretability of the BPI-Facial by calculating the minimum clinically important difference with two anchor-based methods: mean change score and receiver operating curve analysis (Sandhu et al., 2015). The patient global impression of change scale (PGIC) was the anchor used which patients completed on follow-up choosing one of the following options: very much improved, much improved, minimally improved, no change, minimally worse, much worse and very much worse. The mean change score was calculated for one subgroup only ('much improved' $n = 47$) and percentages of change in scores calculated. Patients needed a 30% and 44% improvement in pain intensity worst and average, respectively, to choose the 'much improved option' and a higher percentage change of scores for interference general (54%) and interference facial (63%). Cut-off points were calculated for the domains pain intensity (worse and average), interference general and interference facial for three different models based on the distribution of patients on the PGIC scale – very much improved, much and very much improved, minimally and much and very much improved. The model chosen for analysis was the one which included much and very much improved patients ($n = 159$). Sensitivity and specificity were calculated. For worst and average pain intensity and interference general, sensitivity was 65.5%, 65.7% and 68.3%, respectively, which indicates that there is a moderate percentage of false positive misclassifications. Specificity was higher for all the domains ranging from 71.9% (interference facial) to 90.7% (worst pain intensity).

3.2 | Barrow Neurology Institute Pain Scale

The BNI-PS was used for the first time in a study designed to assess the efficacy of gamma knife radiosurgery in a cohort of TN patients and, according to the authors, it is a pain intensity scale (Rogers et al., 2000); however, requirement for medication is also taken into account. The scoring options outlined on that initial study are: 'I – No trigeminal pain, no medication', 'II – Occasional pain, not requiring medication', 'III – Some pain, adequately controlled with medications', 'IV – Some pain, not adequately controlled with medication' and 'V – Severe pain/no relief'. For a more comprehensive description of this questionnaire, see the review by Sandhu & Lee (2016).

To the best of our knowledge, there is no evidence in the literature of any studies that attest or attempt to validate the BNI-PS for its use in TN studies. The author of the study, where it was used for the first time, was contacted by email for clarification, but we have not obtained a reply. It is not clear if patients complete it or if the data are taken from medical records.

3.2.1 | Interpretability

Two of the included studies were designed to determine the interpretability of the BNI-PS (Reddy et al., 2013, 2014) but the authors of this review agreed that the interpretability of the questionnaire has no significance without evidence of its measurement properties.

3.3 | Visual Analogue Scale

Two of the included studies were designed to determine the interpretability of the VAS (Reddy et al., 2013, 2014). The VAS is a pain intensity scale widely used (Price et al., 1997). It is an unidimensional continuous scale which scores pain intensity on a 10-cm (100 mm) horizontal or vertical line (Kahl & Cleland, 2005). There have been some criticisms of the feasibility of using VAS as a pain intensity measurement, related with the difficulties that the elderly or those with cognitive and physical impairment have in completing it which might result in missing data. For this reason, the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials recommends the numerical rating scale for the assessment of pain intensity in clinical trials of chronic pain (Dworkin et al., 2005).

3.3.1 | Interpretability

The studies which aimed to determine the interpretability of the BNI-PS also set to determine the interpretability of the VAS. Given the absence of VAS psychometric studies within the TN literature, we have not described its interpretability.

3.4 | Penn Facial Pain Scale-Revised

The Penn-FPS-R was developed with the intent to be a revised version of the Penn Facial Pain Scale which was in turn previously called the BPI-Facial (see description above), due to the absence of content validity properties (Symonds et al., 2018). Similar to the BPI-Facial and to the Penn Facial Pain scale, the Penn-FPS-R was designed to capture details on general and TN-specific pain interference. The original BPI-Facial and the Penn Facial Pain Scale included items related to pain interference on activities of daily living specific to patients living with TN such as 'eating a meal', 'touching one's face', 'brushing or flossing one's teeth', 'smiling or laughing', 'talking', 'opening one's mouth widely' and 'eating hard food like apples' (Symonds et al., 2018). These seven items were rated on an 11-point scale (0–10) with anchors 'does not interfere' and 'completely interferes'. In a qualitative study with TN patients (Symonds et al., 2018), the item 'eating hard food like apples' was removed and replaced by 'biting or chewing' and two new items were added, 'self-care' and 'activities affected by temperature changes'. Furthermore, the original BPI-Facial included seven items of general pain interference such as impact of pain on 'walking ability', 'normal work', 'sleep' and 'enjoyment of life', which the participants of this study decided were not relevant in the context of their disease. The finalized version of Penn-FPS-R is a questionnaire which includes 12 items (Symonds et al., 2018).

3.4.1 | Validity

Subscale development and content validity

The subscale development study was of adequate quality (Symonds et al., 2018). In the absence of content validity studies with new patient cohorts, the content validity rating was based on the development study and on the reviewer's ratings which provided moderate quality evidence for sufficient content validity.

3.5 | Trigeminal Neuralgia Quality of Life Score

We have identified one study which aimed to develop a TN-specific QOL subscale for the Quality of Life Instruments for Chronic Diseases (QLICD), which is a questionnaire developed to assess QOL in Chinese populations with chronic diseases (Luo et al., 2019). It consists of a general subscale (QLICD-GM) and disease-specific subscales, which exist for hypertension, irritable bowel syndrome and chronic obstructive pulmonary disease, to name a few (Wan et al., 2011). Despite the use of qualitative methods in the development of the QLICD-GM, patients were not involved nor were they asked about the contents of the questionnaire; therefore, it cannot be assumed that this general subscale has content validity.

3.5.1 | Validity

Subscale development and content validity

One study aimed to develop and confirm content validity for the TN-specific QOL subscale of the QLICD (Luo et al., 2019). The questionnaire development was of doubtful quality, as we could not retrieve any details on the extent of patient involvement. Content validity ratings were based on the development study and on the reviewer's ratings as there was no indication of the subscale being tested on a new cohort. The reviewers deemed content validity

insufficient. This resulted in very low quality for insufficient content validity.

Criterion validity

Criterion validity was described but not assessed in this review, as there are no gold standard questionnaires to assess QOL in TN cohorts.

3.5.2 | Internal structure

Structural validity and internal consistency

The authors used factor analysis to determine the structural validity of the TN QOLS in a study of doubtful quality. Four factors were identified which account for 65.82% of variance. Due to the lack of further information on the factor analysis results (e.g. there was no information on the eigenvalues, nor was there information on the cut-off value for the factor loadings), there is very low evidence for insufficient structure validity.

Cronbach's α was calculated for internal consistency, and results were >0.70 for each of the four factors; however, due very low evidence for insufficient structural validity, internal consistency was deemed indeterminate.

Responsiveness

It is unclear how responsiveness was determined on a study of inadequate quality as there was no evidence of hypothesis testing with a comparator outcome measure. Responsiveness of TN QOLS was insufficient based on very low-quality evidence.

4 | DISCUSSION AND CONCLUSION

This systematic review is the first to use COSMIN guidance to evaluate the measurement properties of PROMs used in patients with TN.

The review identified six studies, in which five different PROMs were used to assess pain intensity, pain interference on activities (general and facial) and pain interference on QOL and daily activities.

A previous systematic review had highlighted the vast number of questionnaires being used in TN studies, with 10 and 9 different questionnaires used for pain relief and pain intensity, respectively (Venda Nova et al., 2020). The results of the present review demonstrated that very few attempts to validate existing questionnaires have been made and that, when it has happened the quality of the evidence has been suboptimal (Table 3). The lack of comparative studies which aim to assess the validity, reproducibility and responsiveness of different questionnaires is striking and has contributed to uncertainties around the best measurement approaches in the TN field.

With the exception of the Penn-FPS-R, which demonstrated moderate evidence for content validity (Symonds et al., 2018), the BPI-Facial demonstrated low evidence for inconsistent content validity (Lee et al., 2010) and the TN QOLS has very low evidence for insufficient content validity (Luo et al., 2019). Content validity is the most important measurement property and involving patients in development studies and validation studies is a requirement according to current guidance (Terwee et al., 2018). Confirming that the

TABLE 3 Evidence synthesis of measurement properties of PROMs used in patients with Trigeminal Neuralgia

	BPI-Facial		Penn FPS – R		TN QOLS	
Measurement property	**Overall Rating**	**Quality of evidence**	**Overall rating**	**Quality of evidence**	**Overall rating**	**Quality of evidence**
Content validity[a]						
Relevance	±	Low	+	Moderate	–	Very low
Comprehensibility	±	Low	+	Moderate	–	Very low
Comprehensiveness	±	Low	+	Moderate	–	Very low
Structural validity	+	Moderate			–	Very low
Internal consistency	+	Moderate				
Measurement invariance						
Reliability						
Measurement error						
Construct validity						
Responsiveness					–	Very low

Note: Empty cells = measurement properties not assessed. "+", sufficient; "−", insufficient; "±", inconsistent.

Abbreviations: BPI-Facial, Brief Pain Inventory – Facial; Penn FPS-R, Penn Facial Pain Scale Revised; TN QOLS, Trigeminal Neuralgia Quality of Life Score.

[a]Construct Validity: rating based on questionnaire development study and reviewers' ratings.

questionnaire is relevant, comprehensible and comprehensive from the patient perspective and for the context of use is at the core of a well-designed patient-reported tool. The questionnaire should be able to capture the patient's experience of living with the disease and how it impacts on their lives (Devlin & Appleby, 2010).

Regulatory agencies like the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) recommend the inclusion of PROs and outcome measures on clinical trials (Gnanasakthy et al., 2019). This is particularly relevant for TN as most studies aim primarily to assess the effectiveness of treatment on pain reduction. In this context, a PRO should be used as a primary outcome/endpoint, given the inherent subjective nature of pain reports. In studies for which no objective primary outcomes exist, the benefit of using methodologically sound patient-reported instruments is even more critical.

The BNI-PS (Rogers et al., 2000) is without a doubt the most replicated outcome measure in surgical studies of TN, yet, no evidence could be found in the literature of any studies which aimed to validate the questionnaire to assess pain intensity in TN cohorts. While it is recognized that the guidance available from COSMIN was not available when the BNI-PS was first developed, it has been widely available for at least a decade. Yet, the BNI-PS continues to be used and its scores are perpetually compared between studies to draw conclusions on treatment effectiveness. This is probably due to its ease of use.

Similar to the BNI-PS, the VAS (Price et al., 1997) has been extensively used in TN literature, with no evidence available for its content validity as illustrated in the present work. A recent systematic review of TN studies found that the VAS has been used not only as a pain intensity outcome measure, as found in 85 of the 193 (44%) studies assessing pain intensity, but, interestingly, in those assessing pain relief, as seen in 18 of the 314 (6%) studies assessing pain relief (Venda Nova et al., 2020). It is possible that it is also due to its ease of use, although it might not be feasible for all patient populations.

The fact that there is a lack of evidence on the content validity of the two most widely used questionnaires for pain intensity and pain relief should be a concern to the field.

The BPI-Facial (Lee et al., 2010) has demonstrated sufficient structural validity and internal consistency in a study of moderate quality, however, as explained above, as it has failed to include patients in its design. As such, these positive results become meaningless in the absence of any evidence to demonstrate content validity. This questionnaire has subsequently been replaced by the Penn-FPS-R (Symonds et al., 2018).

Responsiveness was inadequately assessed for the TN QOLS and no studies assessed it for any other PROMs. Responsiveness is defined by COSMIN as 'the ability of an instrument to detect change over time in the construct to be measured' (Terwee et al., 2011). When designing clinical studies of TN, where the expectation is that the construct under study improves to a certain extent, it is then important to utilize an instrument able to capture the change in scores from baseline to after intervention.

As discussed in the introduction, the importance of using validated questionnaires has been thoroughly described in the literature (Prinsen et al., 2016) and the benefits of doing so, highlighted. Examples of this are the ability to compare study results and draw meaningful conclusions through meta-analysis. Another example relates to the waste of research resources when studies continue to be designed without incorporating psychometrically sound questionnaires. There is no doubt that this is essential for all diseases, but it becomes even more so for rare conditions such as TN where financial support is scarce and recruitment of patients for trials can be challenging (Slade et al., 2018; Zakrzewska et al., 2018). Unusual for the pain field, there are both medical and surgical treatments available for TN with the latter providing more long-term pain relief but with increased risk of complications. In such situations, patients need to be able to compare these when making informed decisions about their treatment.

4.1 | Future directions

The results of this systematic review will inform an ongoing study on the development of a Core Outcome Set for TN (TRINCOS study). However, at this stage, we are unable to make recommendations for the use of any of the questionnaires included in this review, without further psychometric studies.

When designing a study to assess the measurement properties of an instrument, it is important to have in mind the construct or domain of interest (Terwee et al., 2018). We are currently working with TN patients to clarify what outcome domains are important to them. We hypothesize that domains other than pain will be of value to patients, for example, how much interference does the pain cause to their QOL, their daily activities or their mood. Results from a recent cross-sectional study on the burden of illness (O'Callaghan et al., 2020) support this hypothesis; therefore, this information should be taken into account in the design of future psychometric studies. Additionally, TN patients can present with different disease phenotypes, that is, in some, the pain might be purely paroxysmal with variable periods of remission, but others might present with a continuous background pain, which persists in between the attacks (ICOP, 2020). Outcomes of surgical and pharmacological treatment appear to be worse in patients with concomitant pain (Cruccu, 2017; Maarbjerg et al., 2014; Zhang et al., 2013). These distinctive

characteristics of TN should be taken into account when designing or validating questionnaires.

Researchers considering developing and/or validating PROMs could use the results of this review and those of our future work with patients to inform their study designs. We would recommend, as a priority, that instruments currently used in TN are assessed for content validity, and that future studies could include a comparative assessment of multiple instruments. We would also recommend that further psychometric testing is done with the Penn-FPS-R, such as looking at its reliability and responsiveness.

Given the challenge of patient recruitment in TN, researchers working in the field might alternatively want to consider using questionnaires that have been validated in other chronic pain fields and adapt these to be used in cohorts of TN patients, rather than trying to develop a new questionnaire.

4.2 | Limitations

Despite our efforts to conduct an extensive search in five different databases with a validated search filter, we have not looked at grey literature, and we might have left out relevant studies, which could have contributed to the evidence, helping to refute or support our findings. The search strategy was limited to studies in English and it is, therefore, possible that good quality psychometric studies published in other languages were excluded. These limitations might help to explain why we have not been able to make a strong recommendation for the use of any specific questionnaire as we have retrieved very few studies and within these, not all of the nine recommended measurement properties were assessed.

5 | CONCLUSION

This systematic review highlighted the gap that exists in the TN literature on PROMs and has, therefore, failed to provide guidance on which PROMs should be preferred in TN studies. The Penn-FPS-R has promising content validity results, but it needs further investigation of its internal structure and responsiveness.

A substantial future research programme is needed to improve the development and evaluation of PROMs in TN.

CONFLICT OF INTEREST

J.M. Zakrzewska is an author in one of the studies assessed but was not involved in the assessment of the quality of the studies in this review.

AUTHOR CONTRIBUTIONS

CVN conceptualized the study, screened the results, performed the analysis and drafted the manuscript. SRB and JMZ screened the results and critically revised the article providing intellectual contributions to the manuscript. RNR screened the results, performed the analysis and critically revised the article providing intellectual contributions to the manuscript. All authors read and approved the final manuscript.

ORCID

Carolina Venda Nova https://orcid.org/0000-0002-4007-7083

REFERENCES

Cheng, J., Long, J., Hui, X., Lei, D., & Zhang, H. (2017). Effects of microvascular decompression on depression and anxiety in trigeminal neuralgia: A prospective cohort study focused on risk factors and prognosis [Journal Article]. *Clinical Neurology and Neurosurgery*, *161*, 59–64. https://doi.org/10.1016/j.clineuro.2017.08.011

Cleeland, C. S., & Ryan, K. M. (1994). Pain assessment: Global use of the Brief Pain Inventory. *Annals of the Academy of Medicine*, *23*, 129–138.

Cruccu, G. (2017). Trigeminal neuralgia. *CONTINUUM: Lifelong Learning in Neurology*, *23*(2), 396–420. https://doi.org/10.1212/CON.0000000000000451

Cruccu, G., Gronseth, G., Alksne, J., Argoff, C., Brainin, M., Burchiel, K., Nurmikko, T., Zakrzewska, J. M..; American Academy of Neurology, S., & European Federation of Neurological, S. (2008). AAN-EFNS guidelines on trigeminal neuralgia management. *European Journal of Neurology*, *15*(10), 1013–1028. https://doi.org/10.1111/j.1468-1331.2008.02185.x

Devlin, N. J., & Appleby, J. (2010). *Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision-making*. The King's Fund.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., Kerns, R. D., Stucki, G., Allen, R. R., Bellamy, N., Carr, D. B., Chandler, J., Cowan, P., Dionne, R., Galer, B. S., Hertz, S., Jadad, A. R., Kramer, L. D., Manning, D. C., … Witter, J. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, *113*(1), 9–19. https://doi.org/10.1016/j.pain.2004.09.012

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: Scientific advances and future directions. *Psychological Bulletin*, *133*(4), 581–624. https://doi.org/10.1037/0033-2909.133.4.581

Gnanasakthy, A., Barrett, A., Evans, E., D'Alessio, D., & Romano, C. (2019). A review of patient-reported outcomes labeling for oncology drugs approved by the FDA and the EMA (2012–2016). *Value in Health*, *22*(2), 203–209. https://doi.org/10.1016/j.jval.2018.09.2842

ICOP. (2020). International classification of orofacial pain. *Cephalalgia*, *40*(2), 129–221. https://doi.org/10.1177/0333102419893823

IMI. (2018). *Innovative medicines initiative – Big data for better outcomes (BD4BO) – A practical toolkit for the identification, selection and measurement of outcomes including in real-world settings*. Retrieved from http://bd4bo.eu/index.php/toolkit

Kahl, C., & Cleland, J. A. (2005). Visual analogue scale, numeric pain rating scale and the McGill pain Questionnaire: An overview of psychometric properties. *Physical Therapy Reviews*, *10*(2), 123–128. https://doi.org/10.1179/108331905X55776

Kotecha, R., Miller, J. A., Modugula, S., Barnett, G. H., Murphy, E. S., Reddy, C. A., Suh, J. H., Neyman, G., Machado, A., Nagel, S., & Chao, S. T. (2017). Stereotactic radiosurgery for trigeminal neuralgia improves patient-reported quality of life and reduces depression. *International Journal of Radiation Oncology Biology Physics*, *98*(5), 1078–1086. https://doi.org/10.1016/j.ijrobp.2017.04.008

Lee, J. Y., Chen, H. I., Urban, C., Hojat, A., Church, E., Xie, S. X., & Farrar, J. T. (2010). Development of and psychometric testing for the Brief Pain Inventory-Facial in patients with facial pain syndromes. *Journal of Neurosurgery*, *113*(3), 516–523. https://doi.org/10.3171/2010.1.JNS09669

Luo, Y., He, M., Li, C., & Yang, H. (2019). A research on quality of life score (QOLS) of patients with trigeminal neuralgia (TN). *Journal of Infection and Public Health*, *12*(5), 690–694. https://doi.org/10.1016/j.jiph.2019.03.011

Maarbjerg, S., Gozalov, A., Olesen, J., & Bendtsen, L. (2014). Trigeminal neuralgia–a prospective systematic study of clinical characteristics in 158 patients. *Headache*, *54*(10), 1574–1582. https://doi.org/10.1111/head.12441

Magasi, S., Ryan, G., Revicki, D., Lenderking, W., Hays, R. D., Brod, M., Snyder, C., Boers, M., & Cella, D. (2012). Content validity of patient-reported outcome measures: Perspectives from a PROMIS meeting. *Quality of Life Research*, *21*(5), 739–746. https://doi.org/10.1007/s11136-011-9990-8

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mokkink, L. B., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018). *COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) – User manual*. Retrieved from http://www.cosmin.nl/

Mousavi, S. H., Sekula, R. F., Gildengers, A., Gardner, P., & Lunsford, L. D. (2016). Concomitant depression and anxiety negatively affect pain outcomes in surgically managed young patients with trigeminal neuralgia: Long-term clinical outcome. *Surgical Neurology International*, *7*, 98. https://doi.org/10.4103/2152-7806.194145

O'Callaghan, L., Floden, L., Vinikoor-Imler, L., Symonds, T., Giblin, K., Hartford, C., & Zakrzewska, J. M. (2020). Burden of illness of trigeminal neuralgia among patients managed in a specialist center in England. *The Journal of Headache and Pain*, *21*(1), 130. https://doi.org/10.1186/s10194-020-01198-z

Olesen, J. (2018). Headache Classification Committee of the International Headache Society (IHS) the international classification of headache disorders. *Cephalalgia*, *38*(1), 1–211. https://doi.org/10.1177/0333102417738202

Price, D., McGrath, P., Rafii, A., & Buckingham, B. (1997). The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain*, *17*(1), 45–56.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – A practical guideline. *Trials*, *17*.

Reddy, V. K., Parker, S. L., Lockney, D. T., Patrawala, S. A., Su, P. F., & Mericle, R. A. (2014). Percutaneous stereotactic radiofrequency lesioning for trigeminal neuralgia: Determination of minimum clinically important difference in pain improvement for patient-reported outcomes. *Neurosurgery*, *74*(3), 262–266. https://doi.org/10.1227/NEU.0000000000000262

Reddy, V. K., Parker, S. L., Patrawala, S. A., Lockney, D. T., Su, P. F., & Mericle, R. A. (2013). Microvascular decompression for classic trigeminal neuralgia: Determination of minimum clinically important difference in pain improvement for patient reported outcomes. *Neurosurgery*, *72*(5), 749–754. https://doi.org/10.1227/NEU.0b013e318286fad2

Rogers, C. L., Shetter, A. G., Fiedler, J. A., Smith, K. A., Han, P. P., & Speiser, B. L. (2000). Gamma knife radiosurgery for trigeminal neuralgia: The initial experience of The Barrow Neurological Institute. *International Journal of Radiation Oncology Biology Physics*, *47*(4), 1013–1019. https://doi.org/10.1016/s0360-3016(00)00513-7

Sandhu, S. K., Halpern, C. H., Vakhshori, V., Mirsaeedi-Farahani, K., Farrar, J. T., & Lee, J. Y. (2015). Brief pain inventory–facial minimum clinically important difference. *Journal of Neurosurgery*, *122*(1), 180–190. https://doi.org/10.3171/2014.8.Jns132547

Sandhu, S. K., & Lee, J. Y. (2016). Measurement of trigeminal neuralgia pain: Penn Facial Pain Scale. *Neurosurgery Clinics of North America*, *27*(3), 327–336. https://doi.org/10.1016/j.nec.2016.02.003

Slade, A., Isa, F., Kyte, D., Pankhurst, T., Kerecuk, L., Ferguson, J., Lipkin, G., & Calvert, M. (2018). Patient reported outcome measures in rare diseases: A narrative review. *Orphanet Journal of Rare Diseases*, *13*(1), 61. https://doi.org/10.1186/s13023-018-0810-x

Sullivan, M. D., & Ballantyne, J. C. (2016). Must we reduce pain intensity to treat chronic pain? *Pain*, *157*(1), 65–69. https://doi.org/10.1097/j.pain.0000000000000336

Symonds, T., Randall, J. A., Hoffman, D., Zakrzewska, J., Gehringer, W., & Lee, J. Y. K. (2016). Content validity of the Penn facial pain scale: Cognitive debriefing in patients with trigeminal neuralgia. *Postgraduate Medicine*, *128*(Suppl. 2), 87–88. https://doi.org/10.1080/00325481.2016.1224633

Symonds, T., Randall, J. A., Hoffman, D. L., Zakrzewska, J. M., Gehringer, W., & Lee, J. Y. (2018). Measuring the impact of trigeminal neuralgia pain: The Penn Facial Pain Scale-Revised. *Journal of Pain Research*, *11*, 1067–1073. https://doi.org/10.2147/jpr.S152958

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*(8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5

Terwee, C. B., Knol, D. L., de Vet, H. C. W., & Mokkink, L. B. (2011). Responsiveness. *Measurement in medicine: A practical guide (practical guides to biostatistics and epidemiology)* (pp. 202–226). Cambridge University Press. https://doi.org/10.1017/CBO9780511996214.011

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170. https://doi.org/10.1007/s11136-018-1829-0

Venda Nova, C., Zakrzewska, J. M., Baker, S. R., & Ni Riordain, R. (2020). Treatment outcomes in trigeminal neuralgia – A systematic review of domains, dimensions and measures. *World Neurosurgery: X*, *6*, 100070. https://doi.org/10.1016/j.wnsx.2020.100070

Wan, C., Tu, X., Messing, S., Li, X., Yang, Z., Zhao, X., Gao, L. I., Yang, Y., Pan, J., & Zhou, Z. (2011). Development and validation of

the general module of the system of quality of life instruments for chronic diseases and its comparison with SF-36. *Journal of Pain and Symptom Management*, *42*(1), 93–104. https://doi.org/10.1016/j.jpainsymman.2010.09.024

Weldring, T., & Smith, S. M. S. (2013). Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Services Insights*, *6*, 61–68. https://doi.org/10.4137/HSI.S11093

Zakrzewska, J. M., Palmer, J., Bendtsen, L., Di Stefano, G., Ettlin, D. A., Maarbjerg, S., Obermann, M., Morisset, V., Steiner, D., Tate, S., & Cruccu, G. (2018). Challenges recruiting to a proof-of-concept pharmaceutical trial for a rare disease: The trigeminal neuralgia experience. *Trials*, *19*(1), 704. https://doi.org/10.1186/s13063-018-3045-1

Zakrzewska, J. M., & Relton, C. (2016). Future directions for surgical trial designs in trigeminal neuralgia. *Neurosurgery Clinics of North America*, *27*(3), 353–363. https://doi.org/10.1016/j.nec.2016.02.011

Zhang, H., Lei, D., You, C., Mao, B.-Y., Wu, B., & Fang, Y. (2013). The long-term outcome predictors of pure microvascular decompression for primary trigeminal neuralgia. *World Neurosurgery*, *79*(5–6), 756–762. https://doi.org/10.1016/j.wneu.2012.01.040

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Venda Nova C, Zakrzewska JM, Ni Riordain R, Baker S. Patient reported outcome measures in trigeminal neuralgia – A systematic review of psychometric performance. *Eur J Pain*. 2021;00:1–13. https://doi.org/10.1002/ejp.1779