GMR
The Original

MINIMAL CLINICALLY IMPORTANT DIFFERENCE AND CLINICAL INTERPRETABILITY OF THE CERVICOGENIC HEADACHE SEVERITY QUESTIONNAIRE (CEH-SEQ)

Prachi Bakul Choksi[1]*, Ankurkumar Premjibhai Khant[2]

[1]PhD Scholar, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat, India
[2]Associate Professor, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat, India
*Correspondence: prachichoksi88@gmail.com

ABSTRACT

Background: The Cervicogenic Headache Severity Questionnaire (CeH-SeQ) is a novel patient-reported outcome measure (PROM) developed to screen for and grade the severity of cervicogenic headache (CeH). Its content validity and test–retest reliability have previously been established, but the minimal clinically important difference (MCID) and the clinical interpretability of its severity bands had not yet been empirically defined.

Objective: To derive an anchor-based and distribution-based MCID for the CeH-SeQ Part 2 severity score, and to evaluate the clinical interpretability of its predefined severity bands, using data from a preliminary prospective cohort.

Methods: Thirty adults with clinically confirmed CeH (CeH-SeQ Part 1 score ≥ 8) completed the CeH-SeQ, Patient Global Impression of Change (PGIC), Numerical Pain Rating Scale (NPRS), Neck Disability Index (NDI), and Headache Impact Test-6 (HIT-6) at baseline and at 4-week follow-up. Anchor-based MCID was derived via receiver operating characteristic (ROC)/Youden Index analysis using PGIC-defined 'minimally improved' status as the external anchor. Distribution-based estimates (half standard deviation [SD], standard error of measurement [SEM], smallest detectable change [SDC]) were calculated from a co-located 72-hour test–retest reliability sub-study (n = 30, ICC(2,1) model). Agreement between anchor- and CeH-SeQ-defined improvement was quantified with Cohen's kappa, and concurrent responsiveness was assessed with Spearman correlations against NPRS, NDI, and HIT-6 change scores.

Results: Mean baseline CeH-SeQ severity was 76.48% ± 13.96% (Very Severe band in 20/30 participants), improving to 74.33% ± 14.11% at 4 weeks (mean change 2.14 ± 1.21 percentage points). The anchor-based MCID was 2.0 percentage points (sensitivity 0.75, specificity 0.50, Youden J = 0.25, AUC = 0.61). Distribution-based estimates were higher than the anchor-based value: half-SD = 6.98%, SEM = 0.88%, SDC = 2.44%. Test–retest reliability was excellent (ICC(2,1) = 0.996). Agreement between PGIC- and CeH-SeQ-defined improvement was fair ($\kappa = 0.25$). Correlations between CeH-SeQ change and NPRS ($\rho = 0.11$, $p = 0.577$), NDI ($\rho = 0.33$, $p = 0.071$), and HIT-6 change ($\rho = 0.32$, $p = 0.081$) were weak to moderate and did not reach statistical significance.

Conclusion: In this preliminary cohort, the empirically derived anchor-based MCID for the CeH-SeQ (≈ 2 percentage points) was substantially smaller than the a priori projected range of 10–15%, and concurrent responsiveness against established pain and disability measures was weaker than hypothesised, despite excellent test–retest reliability. These findings support the CeH-SeQ's measurement precision but indicate that its MCID and severity-band thresholds require confirmation in the full-scale, adequately powered validation cohort before adoption as a definitive clinical or trial-endpoint threshold.

KEYWORDS: cervicogenic headache; minimal clinically important difference; patient-reported outcome measure; CeH-SeQ; clinical interpretability; MCID; neck pain; headache severity

1. INTRODUCTION

Cervicogenic headache (CeH) is a secondary headache disorder arising from structural or functional pathology of the cervical spine and is estimated to account for a meaningful proportion of chronic headache presentations encountered in musculoskeletal and physiotherapy practice. It remains diagnostically challenging, frequently overlapping in presentation with migraine and tension-type headache.

The CeH-SeQ was developed to address the absence of a CeH-specific, dual-function patient-reported outcome measure, combining a 23-item dichotomous diagnostic screen (Part 1, aligned with ICHD-3 and CHISG criteria) with a seven-domain severity scale (Part 2, maximum raw score 36, expressed as a percentage). While the instrument's content validity (S-CVI/Ave = 0.82) and test–retest reliability (ICC = 0.85, Pearson r = 0.91) have previously been reported, its clinical interpretability — specifically its minimal clinically important difference (MCID) and the empirical validity of its predefined severity bands (Mild, Moderate, Severe, Very Severe) — had not been established.

This paper reports the results of the first empirical MCID and clinical-interpretability analysis of the CeH-SeQ, conducted in a preliminary prospective cohort as the initial phase of a planned larger validation study. Anchor-based and distribution-based methods were triangulated, following COSMIN-consistent methodology, using the Patient Global Impression of

Change (PGIC) as the primary anchor and the Numerical Pain Rating Scale (NPRS), Neck Disability Index (NDI), and Headache Impact Test-6 (HIT-6) as secondary anchors and concurrent responsiveness comparators.

2. METHODS

2.1 Design and Participants

A prospective, single-cohort, repeated-measures design was used. Thirty adults with a primary complaint of head and/or neck pain, meeting CeH-SeQ Part 1 screening criteria (score ≥ 8, indicating likely CeH) and free of the instrument's stated exclusion criteria (recent trauma, unrelated neurological symptoms, cervical instability, myelopathy, occipital neuralgia, or intracranial/vascular pathology), were consecutively enrolled at a physiotherapy outpatient setting affiliated with Marwadi University, Rajkot. This preliminary cohort (n = 30) represents the initial analysis phase of a larger planned study (target n = 144, based on a priori ROC power calculation); results below should therefore be interpreted as pilot estimates pending confirmation in the full sample.

2.2 Outcome Measures and Procedure

Participants completed the CeH-SeQ Part 2 severity scale, NPRS, NDI, and HIT-6 at baseline and again at 4-week follow-up, alongside the 7-point PGIC at follow-up. Participants rating PGIC ≥ +1 ('a little better' or greater) were classified as the 'minimally improved' anchor group. A separate, co-located test–retest reliability sub-study was conducted in the same 30 clinically stable participants, who completed the CeH-SeQ twice, 72 hours apart, without intervening treatment.

2.3 Statistical Analysis

Anchor-based MCID was estimated by ROC curve analysis, identifying the CeH-SeQ change-score threshold that maximised the Youden Index (sensitivity + specificity − 1) for classifying PGIC-defined minimal improvement. Distribution-based estimates comprised half the baseline standard deviation (0.5 SD) and the standard error of measurement (SEM = SD$\sqrt{}$[1−ICC]), with ICC(2,1) computed from the test–retest sub-study using the one-way ANOVA mean-square method; the smallest detectable change (SDC = SEM × 1.96 × $\sqrt{}$2) was derived from the SEM. Agreement between PGIC-defined improvement and CeH-SeQ MCID-threshold-defined improvement was quantified with Cohen's kappa. Concurrent responsiveness was evaluated using Spearman's rank correlations between CeH-SeQ change scores and NPRS, NDI, and HIT-6 change scores. Analyses were performed in Python (pandas/SciPy/scikit-learn) with cross-validation against the source calculation workbook; significance was set at $p < 0.05$.

3. RESULTS

3.1 Participant Characteristics

Thirty participants were analysed (17 female, 13 male; mean age 35.57 ± 10.56 years). At baseline, CeH-SeQ severity band distribution was: Mild, 0 (0%); Moderate, 2 (6.7%); Severe, 8 (26.7%); Very Severe, 20 (66.7%). Mean baseline CeH-SeQ score was 76.48% ± 13.96%, falling at follow-up to 74.33% ± 14.11% — a mean absolute change of 2.14 ± 1.21 percentage points. Twenty of 30 participants (66.7%) met PGIC criteria for 'minimally improved' status; the remaining 10 (33.3%) reported no meaningful change.

Table 1. Participant and Baseline Characteristics (n = 30)

Variable	Value
Age (years), mean ± SD	35.57 ± 10.56
Sex, female / male	17 / 13
Baseline CeH-SeQ (%), mean ± SD	76.48 ± 13.96
4-week CeH-SeQ (%), mean ± SD	74.33 ± 14.11
Baseline band: Moderate / Severe / Very Severe	2 / 8 / 20
PGIC 'minimally improved', n (%)	20 (66.7)

3.2 Anchor-Based MCID (ROC/Youden Index)

The Youden Index was maximised at a CeH-SeQ change-score threshold of 2.0 percentage points, at which sensitivity for classifying PGIC-defined minimal improvement was 0.75 and specificity was 0.50 (Youden J = 0.25). The area under the ROC curve was 0.61, indicating modest discrimination. This anchor-based MCID estimate is substantially smaller than the 10–15% range hypothesised a priori by analogy with comparator PROMs.

Table 2. Anchor-Based MCID — ROC/Youden Index Results

Metric	Value
Optimal MCID threshold (Δ CeH-SeQ, %)	2.0
Sensitivity	0.75
Specificity	0.50

Youden Index (J)	0.25
Area under ROC curve (AUC)	0.61
N, minimally improved / not improved	20 / 10

3.3 Distribution-Based Estimates

Distribution-based estimates, derived from the baseline score spread and the 72-hour test–retest sub-study, yielded consistently higher values than the anchor-based MCID. Test–retest reliability was excellent (ICC(2,1) = 0.996), producing a small SEM (0.88%) and SDC (2.44%); the half-SD estimate (6.98%) was more than three times the anchor-based threshold.

Table 3. Distribution-Based Estimates

Metric	Value
Baseline SD (CeH-SeQ, %)	13.96
Half-SD (0.5 × SD)	6.98%
ICC(2,1), test–retest (n = 30, 72 hr)	0.996
SEM (= SD√[1−ICC])	0.88%
SDC (= SEM × 1.96 × √2)	2.44%

3.4 Agreement Between Anchor- and CeH-SeQ-Defined Improvement

Classifying participants as 'improved' on the CeH-SeQ using the empirically derived 2.0-percentage-point threshold and cross-tabulating this against PGIC-defined improvement yielded 15 concordant 'improved' classifications, 5 concordant 'not improved' classifications, and 10 discordant classifications (5 false positive, 5 false negative), corresponding to fair agreement (Cohen's κ = 0.25; Landis & Koch benchmark: 0.21–0.40).

Table 4. Agreement Between PGIC-Defined and CeH-SeQ MCID-Threshold-Defined Improvement

	CeH-SeQ: Not Improved	CeH-SeQ: Improved	Row Total
PGIC: Not improved	5	5	10
PGIC: Minimally improved	5	15	20
Column Total	10	20	30

3.5 Concurrent Responsiveness

Secondary outcome measures showed the expected direction of change over the 4-week interval: mean NPRS fell from 6.97 ± 1.61 to 5.33 ± 1.63 (Δ1.63 ± 1.54); mean NDI fell from 43.43% ± 7.28% to 29.83% ± 7.12% (Δ13.60% ± 5.22%); and mean HIT-6 fell from 65.67 ± 3.74 to 56.00 ± 4.39 (Δ9.67 ± 1.73). However, Spearman correlations between CeH-SeQ change scores and these comparator change scores were weak to moderate and did not reach conventional statistical significance in this sample size: NPRS (ρ = 0.11, p = 0.577), NDI (ρ = 0.33, p = 0.071), HIT-6 (ρ = 0.32, p = 0.081) — below the moderate-to-strong correlations (r = 0.55–0.70) hypothesised a priori.

Table 5. Spearman Correlations — CeH-SeQ Change vs. Comparator Change Scores (n = 30)

Comparison	Spearman's ρ	p-value	Interpretation
CeH-SeQ Δ vs NPRS Δ	0.11	0.577	Weak, NS
CeH-SeQ Δ vs NDI Δ	0.33	0.071	Weak–moderate, NS
CeH-SeQ Δ vs HIT-6 Δ	0.32	0.081	Weak–moderate, NS

3.6 Observed vs. Projected MCID

Table 6. Observed CeH-SeQ MCID in Context of A Priori Projection and Comparator PROMs

Instrument	Scale Range	MCID	Method
HIT-6	36–78	2.3 points	Anchor-based
MIDAS	0–21+	5.1 points	Anchor-based
NDI	0–100%	7.0–10.5%	Anchor/Distribution

NPRS (neck pain)	0–10	2.0 points	Anchor-based
CeH-SeQ (projected, protocol)	0–100%	10–15%	ROC + 0.5 SD + SEM (hypothesised)
CeH-SeQ (observed, this study)	0–100%	2.0% (anchor); 6.98% (half-SD)	ROC/Youden; 0.5 SD

4. DISCUSSION

This preliminary analysis provides the first empirical MCID estimates for the CeH-SeQ. Three findings merit particular attention. First, the anchor-based MCID (2.0 percentage points) was markedly smaller than the 10–15% range hypothesised in the study protocol by analogy with the HIT-6, MIDAS, and NDI. Second, the anchor-based and distribution-based estimates diverged substantially — the half-SD estimate (6.98%) was more than three times the anchor-based value — whereas MCID methodology typically expects reasonable convergence between these approaches when an instrument's clinically important change aligns well with its measurement variability. Third, concurrent correlations between CeH-SeQ change and established pain, disability, and headache-impact measures were weaker than anticipated and did not reach statistical significance, despite all instruments changing in the expected clinically favourable direction.

Taken together, these results should be interpreted cautiously rather than as a definitive MCID for clinical adoption. The modest ROC discrimination (AUC = 0.61) and fair κ agreement (0.25) between PGIC- and CeH-SeQ-defined improvement suggest that, in this sample, the CeH-SeQ percentage-change score alone imperfectly captured patient-perceived global improvement. Plausible contributors include the small sample size (n = 30 versus the planned 144, limiting ROC and correlation precision), the narrow 4-week observation window, potential floor/ceiling effects given that 93% of participants fell in the Severe or Very Severe bands at baseline, and possible misclassification within the PGIC anchor itself. It is also possible that the seven-domain structure of CeH-SeQ Part 2, spanning frequency, intensity, duration, interference, associated symptoms, treatment effectiveness, and life impact, is responsive across domains that do not move in lock-step with a single global change rating, diluting the anchor-based signal even where genuine improvement occurred.

By contrast, the distribution-based indices were highly favourable: an ICC(2,1) of 0.996 across a 72-hour retest interval indicates excellent measurement stability, and the correspondingly small SEM (0.88%) and SDC (2.44%) indicate that the instrument can detect small true changes with low measurement error. This combination — excellent reliability alongside an unexpectedly small and imprecisely discriminating anchor-based MCID — is consistent with an instrument that is measuring consistently but whose clinically important change threshold has not yet been pinned down with adequate anchor-group separation. A larger, adequately powered cohort, incorporating a broader range of baseline severity (including Mild and Moderate presentations, underrepresented here), a longer follow-up interval to allow fuller expression of treatment-related change, and a dichotomised PGIC anchor free of the coding irregularities noted during data cleaning, would be expected to narrow this gap and yield a more stable, generalisable MCID estimate.

These findings should be read alongside the CeH-SeQ's previously reported psychometric profile — satisfactory content validity (S-CVI/Ave = 0.82) and strong test–retest reliability (ICC = 0.85, κ = 0.98) in an independent validation cohort — which remains unaffected by the present results. The present study addresses a distinct psychometric property, interpretability, that is necessary before the CeH-SeQ can be confidently used to judge whether an individual patient's or trial arm's change in score represents a clinically meaningful benefit rather than measurement noise or the natural fluctuation of CeH symptoms.

5. LIMITATIONS

This analysis is constrained by its preliminary sample size (n = 30 versus the protocol-specified target of 144), which limits the precision of the ROC-derived MCID, the stability of the Youden Index, and the power to detect the hypothesised concurrent correlations. The cohort was drawn from a single physiotherapy centre and was heavily weighted toward Severe and Very Severe baseline presentations, limiting generalisability across the full severity spectrum and constraining evaluation of the Mild and Moderate severity bands. The 4-week follow-up window, while pragmatic, may not have been sufficient to capture the full trajectory of clinically important change for all treatment modalities. Data-quality irregularities were identified in the underlying PGIC coding for a subset of participants during analysis; results reported here rely on the validated dichotomous ‘minimally improved’ classification rather than the raw ordinal PGIC code, which should be re-verified in subsequent data collection. Finally, reliance on a single primary anchor (PGIC), although methodologically standard, may not capture all dimensions of clinically meaningful change relevant to CeH.

6. CONCLUSION

In this preliminary cohort, the CeH-SeQ demonstrated excellent test–retest reliability but an anchor-based MCID (≈ 2 percentage points) considerably smaller than hypothesised, alongside weaker-than-expected agreement with PGIC-defined improvement and concurrent measures of pain, disability, and headache impact. These results represent an initial, hypothesis-generating step toward establishing the CeH-SeQ's clinical interpretability rather than a definitive threshold for practice or trial design. Completion of the full-scale validation cohort (target n = 144), with broader baseline severity representation and a verified PGIC anchor, is required before the MCID and severity-band thresholds reported here are adopted for clinical decision-making or used to power future CeH-SeQ-based intervention trials.

Declarations

Ethics Approval: This preliminary analysis was conducted under the study protocol submitted to the Marwadi University Institutional Ethics Committee.

Informed Consent: Written informed consent was obtained from all participants.

Funding: No external funding was received for this study.

Conflict of Interest: The authors declare no conflicts of interest.

Author Contributions: PBC: instrument development, methodology, data analysis, manuscript preparation. APK: statistical design, data interpretation, manuscript review and editing. Both authors approved the final version.

Data Availability: De-identified study data are available from the corresponding author upon reasonable request.

REFERENCES

1. Sjaastad O, Fredriksen TA, Pfaffenrath V. Cervicogenic headache: diagnostic criteria. The Cervicogenic Headache International Study Group. Headache. 1998;38(6):442-445.
2. Biondi DM. Cervicogenic headache: a review of diagnostic and treatment strategies. J Am Osteopath Assoc. 2005;105(4 Suppl 2):16S-22S.
3. Haldeman S, Dagenais S. Cervicogenic headaches: a critical review. Spine J. 2001;1(1):31-46.
4. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition. Cephalalgia. 2018;38(1):1-211.
5. Sjaastad O, Fredriksen TA, Pfaffenrath V. Cervicogenic headache: diagnostic criteria. Headache. 1990;30(11):725-726.
6. Jadeja ND, Choksi PB, Khant AP. CeH-SeQ: Cervicogenic Headache Severity – A Patient-Reported Assessment Tool. Marwadi University; 2024.
7. Choksi PB, Khant AP. Validation and Test–Retest Reliability of the Cervicogenic Headache Severity Questionnaire (CeH-SeQ): A Patient-Reported Assessment Tool. EPJ Web Conf. 2026;348:01004.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-415.
9. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77(4):371-383.
10. Revicki DA, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61(2):102-109.
11. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther. 1991;14(7):409-415.
12. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. Sports Med. 2010;40(7):525-537.
13. Fernandez-de-las-Penas C, Alonso-Blanco C, Cuadrado ML, Pareja JA. Forward head posture as a risk factor for neck pain: prospective study. Cephalalgia. 2006;26(3):314-319.
14. Jull G, Trott P, Potter H, et al. A randomized controlled trial of exercise and manipulative therapy for cervicogenic headache. Spine (Phila Pa 1976). 2002;27(17):1835-1843.
15. Bogduk N. The neck and headaches. Neurol Clin. 2004;22(1):151-171.
16. Kosinski M, Bayliss MS, Bjorner JB, et al. A six-item short-form survey for measuring headache impact: the HIT-6. Qual Life Res. 2003;12(8):963-974.
17. Stewart WF, Lipton RB, Dowson AJ, Sawyer J. Development and testing of the Migraine Disability Assessment (MIDAS) Questionnaire to assess headache-related disability. Neurology. 2001;56(6 Suppl 1):S20-S28.
18. Rendas-Baum R, Yang M, Varon SF, Bloudek LM, DeGryse RE, Kosinski M. Validation of the Headache Impact Test (HIT-6) in patients with chronic migraine. Health Qual Life Outcomes. 2014;12:117.
19. Pool JJ, Ostelo RW, Hoving JL, Bouter LM, de Vet HC. Minimal clinically important change of the Neck Disability Index and the Numerical Rating Scale for patients with neck pain. Spine (Phila Pa 1976). 2007;32(26):3047-3051.
20. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41(5):582-592.
21. Dworkin RH, Turk DC, McDermott MP, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain. 2009;146(3):238-244.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.
23. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. Biom J. 2005;47(4):458-472.
24. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-35.