*Iran J Public Health, Vol. 55, No.1, Jan 2026, pp.1-13*

**Review Article**

# Internal Consistency Reliability of Patient-Reported Outcomes Measurement Information System Measures Used in Symptom Cluster Research: A Systematic Review and Reliability Generalization Meta-Analysis

***Jeehye Jun*** [1], ****Choa Sung*** [2], ***Chang Gi Park*** [3]

1. *Red Cross College of Nursing, Chung-Ang University, Seoul, Republic of Korea*
2. *Biobehavioral Nursing Science, College of Nursing, University of Illinois Chicago, Chicago, USA*
3. *Department of Population Health Nursing Science, College of Nursing, University of Illinois Chicago, Chicago, USA*

***Corresponding Author:** Email: cathy16234@gmail.com

**(Received 05 Jul 2025; accepted 24 Oct 2025)**

**Abstract**

**Background:** Patients with acute and chronic conditions often experience multiple symptoms known as symptom clusters. The Patient-Reported Outcomes Measurement Information System (PROMIS) is widely used to assess health status across various conditions, but its suitability for identifying symptom clusters remains unclear. Therefore, we examined the internal consistency reliability of PROMIS tools used to measure symptom clusters in adults through a systematic review and a reliability generalization meta-analysis.

**Methods:** We searched four electronic databases (PubMed, CINAHL, ProQuest, and Embase) for relevant articles published through December 31, 2024, including studies that measured symptom clusters in adults using at least one PROMIS measure. Meta-regression using a random effects model was performed to assess study heterogeneity, and funnel plots were employed to evaluate publication bias.

**Results:** The systematic review included 24 studies of 27,982 subjects with or without diseases in community, inpatient, and outpatient settings. Twenty PROMIS domains were used for symptom cluster research, and anxiety and depression were the most frequently used domains. In our reliability generalization meta-analysis of four studies, Cronbach's alpha coefficients indicated good internal consistency reliability across five PROMIS domains (anxiety, depression, fatigue, pain, and sleep disturbance), with an average reliability of 0.91.

**Conclusion:** PROMIS measures may be reliable for assessing symptom clusters in adults and could serve as valuable tools for researchers and clinicians in patient assessment and symptom management. Nevertheless, future research should rigorously examine the reliability and validity of PROMIS tools in this context.

Trial Registration Information: CRD42022373953 (PROSPERO).

**Keywords:** Symptom management; Internal consistency reliability; Systematic review; Meta-analyses

## Introduction

Patients with acute and chronic conditions experience a multitude of symptoms across the disease trajectory and within different populations. When symptoms overlap or exist synergistically,


DOI: https://doi.org/10.18502/ijph.v55i1.20968

they are collectively known as a symptom cluster (1). Managing one symptom may have a crossover effect on other symptoms within a cluster that shares an underlying mechanism (2). Conceptually, two approaches are currently being applied to assess symptom clusters: the variable-centered (i.e., *de novo*) and person-centered (i.e., *a priori*) approaches (3). A variable-centered approach clusters symptoms using factor, cluster, and network analyses, while a person-centered approach clusters subgroups of patients using latent class and profile analyses (4). By means of these advanced statistical methods, a wide variety of symptom clusters have been identified over the past two decades. However, the use of different symptom sets has restricted researchers' ability to generalize their findings across conditions and populations.

The Patient-Reported Outcomes Measurement Information System (PROMIS) measures address this variability through standardized measurement tools. PROMIS was developed to ensure the standardization and interoperability of patient-reported outcome measurement (4, 5). Through rigorous development methods and direct input from clinicians, researchers, and patients, PROMIS measures capture important domains of health and functioning (6, 7) and are widely used to monitor various aspects of an individual's health status, including their physical, mental, and social well-being (8, 9). Importantly, PROMIS provides disease-agnostic score metrics (i.e., scores that are comparable across diagnoses) and covers 75% of the universal patient-reported outcome sets created by the International Consortium for Health Outcomes (10, 11). Therefore, by utilizing PROMIS tools, researchers can obtain a holistic view of symptom experiences and can potentially identify their interrelationships within a symptom cluster (12). Furthermore, the use of PROMIS has increased in appraising symptoms in chronic conditions, with studies demonstrating its feasibility across diverse populations (13, 14). These characteristics position PROMIS as a strong candidate for a core symptom measurement in symptom cluster studies.

Due to their stringent development and validation, PROMIS measures have demonstrated strong psychometric properties, including reliability, validity, and responsiveness, providing researchers with confidence in the accuracy and precision of the data collected with these tools (15). However, despite extensive evidence supporting their use in diverse populations, it remains unclear whether PROMIS instruments are appropriately suited for identifying symptom clusters, a more complex construct involving the co-occurrence of multiple symptoms.

Therefore, the aim of this systematic review was to synthesize the literature on PROMIS instruments employed for adults in symptom cluster research, examining their application, methodological quality, and internal consistency reliability where reported.

## Methods

This systematic review and meta-analysis study was registered in PROSPERO. For the literature search, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was followed to select relevant articles. Cooper's approach was employed to conduct the research synthesis and meta-analysis (16).

### *Search Methods*

Four electronic databases (MEDLINE/PubMed, CINAHL, ProQuest, and Embase) were used to find relevant articles based on the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) methodology. COSMIN provides database-tailored search filters that can be used to properly identify all studies addressing measurement properties for systematic reviews of patient-reported outcome measures (17, 18). Based on the COSMIN methodology and with assistance from a reference librarian (R.R.), specific search terms were generated for each database. The core search strategy combined "symptom cluster" and its variations (e.g., symptom* OR syndrome*) with "PROMIS" and its full name "patient-reported outcomes

measurement information system," adapted to each database's syntax and controlled vocabulary. The complete search strings used for each database are provided in Appendix 1 (Not published). In addition, the reference lists of the included studies were reviewed by the authors (J.J. and C.S.) in an attempt to identify additional sources.

### *Inclusion and Exclusion Criteria*

For the systematic review, studies were included if they 1) involved symptom cluster research with adults (aged 18 years or older), 2) used at least one PROMIS measure (any version or domain), and 3) were published in English by December 31, 2024. In this article, a symptom cluster is operationalized as two or more co-occurring symptoms identified based on either of the two conceptual approaches (*de novo* or *a priori*) using advanced statistical methods (4). Studies were excluded if they 1) were grey literature (e.g., a dissertation, report, or conference abstract) or 2) used translated versions of PROMIS measures. Studies selected for the systematic review were included in the meta-analysis if they provided at least one internal consistency reliability value for a PROMIS measure. If a study reported a range of values, the corresponding author was contacted via email to obtain specific reliability values.

### *Search Outcome*

A PRISMA flowchart summarizing the study selection process is depicted in Fig. 1. The database search identified 15,541 studies.

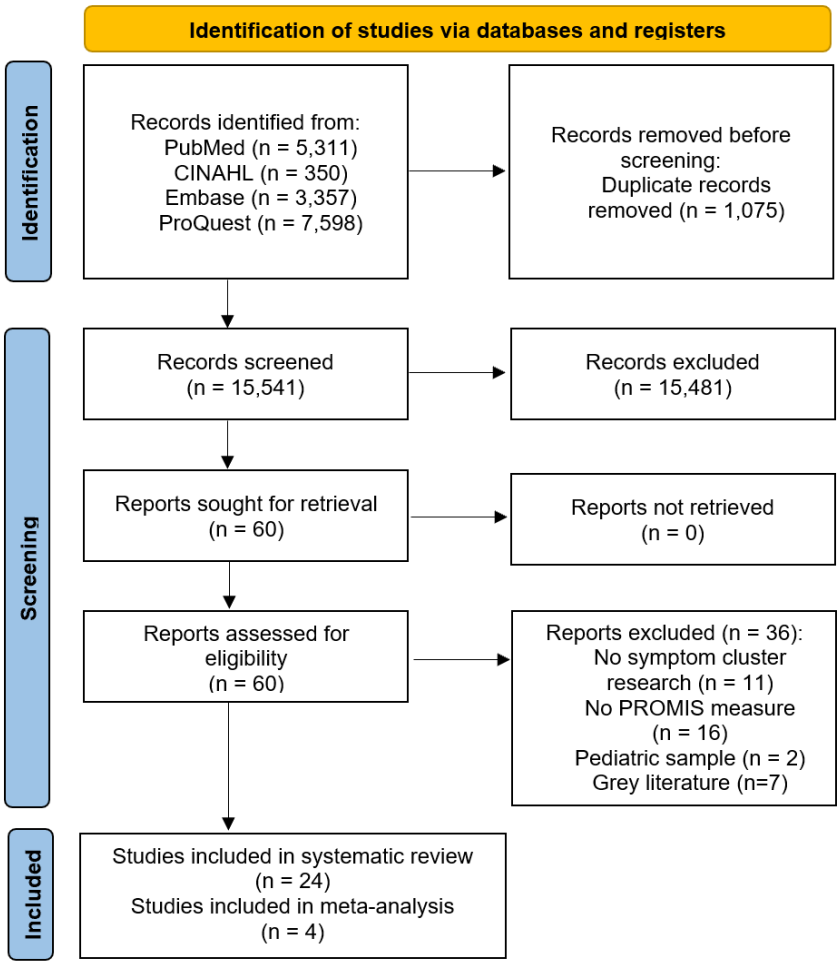


**Fig. 1: PRISMA flow diagram.** *Note.* PROMIS=Patient-Reported Outcomes Measurement Information System

The screening process was conducted in two sequential stages. In stage 1, the two reviewers (J.J. and C.S.) independently screened titles and abstracts, removing duplicates and studies clearly unrelated to symptom clusters or patient-reported outcomes. For the remaining studies, abstracts were reviewed to identify potential eligibility based on key criteria. In stage 2, full-text articles were obtained for the 60 studies that passed initial screening. During full-text review, reviewers systematically evaluated each article by first confirming symptom cluster analysis in the methods section, then verifying PROMIS use in the measures/instruments section, and finally checking participant characteristics to confirm adult populations. Based on the inclusion criteria, 24 studies were chosen for the systematic review. Review of the reference lists of the 24 studies revealed no additional articles to be included.

Among the 24 studies, four reported Cronbach's alpha values for internal consistency reliability (19-22), and no study reported reliability information other than Cronbach's alpha values. For two studies that provided only ranges of Cronbach's alpha coefficients for multiple PROMIS domains, the corresponding authors were contacted to obtain an individual value for each domain. Because those authors did not provide the values requested, the two studies were excluded from the meta-analysis. Therefore, the four studies were included in a reliability generalization meta-analysis.

### *Quality Appraisal*

The quality of the included studies was assessed using the revised tool for Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2). While QUADAS-2 was designed for diagnostic accuracy studies, it has been adapted and used to appraise validation studies of outcome measures in several systematic reviews (23, 24). This tool consists of four domains posing risk of bias (patient selection, index test, reference standard, and study flow and timing of index tests) and three domains of applicability concerns (patient selection, index test, and reference standard) (25). QUADAS-2 has been widely used to evaluate whether a study's context introduced bias in a given risk of bias domain and whether the study's research question aligned with each applicability concern domain. The seven domains are rated as "low," "high," or "unclear" when related data are insufficiently reported in a study. For the current review, two authors (J.J. and C.S.) independently evaluated each study in terms of the seven domains and reached a quality assessment consensus when any differences of opinion arose.

### *Data Abstraction*

Covidence, a web-based systematic review software program, was used for study selection. Two reviewers (J.J. and C.S.) independently screened retrieved titles and abstracts and performed full-text reviews to identify eligible studies. When differences of opinion arose between the reviewers, deliberate discussions were held until they reached a consensus on the studies to be included. The software computed proportionate agreement as the percentage of studies where both reviewers made concordant inclusion/exclusion decisions. The proportionate agreement between the reviewers was 99.7% and 90.7% for the title and abstract review and full-text review, respectively, indicating good inter-rater reliability.

The following information was extracted from the studies included for the systematic review: author(s), publication year, country, study setting (inpatient, outpatient, or community), study design, characteristics of the study population (sample size, age, gender), statistical methods used for identification of symptom clusters, types and versions of PROMIS measures used, reliability values, and main findings.

### *Synthesis*

The meta-analysis was performed using Stata version 16.1. A *P*-value less than 0.05 was used to represent statistical significance. The following steps were applied to calculate overall mean values for internal consistency reliability. The effect size statistic was derived from Cronbach's alpha

reliability coefficient for PROMIS scores in each study. Second, the standard error (SE) was calculated for each study based on the reported number of samples (n) and Cronbach's alpha values ($r$):

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Third, meta-regression using a random effects model was performed by including all PROMIS domains to obtain overall mean values for internal consistency reliability. Moreover, meta-regression of each PROMIS domain (e.g., pain or depression) was performed to obtain overall mean reliability values for that domain.

Forest plots of Cronbach's alpha values were generated for each meta-regression, and the $I^2$ index was calculated to determine the extent of effect size heterogeneity across studies. For example, $I^2 = 50$ indicates that half the total variability among effect sizes is due to study heterogeneity, and percentages of 25, 50, and 75 indicate low, medium, and high heterogeneity, respectively (26). To assess publication bias, funnel plots were first generated to visualize whether or not the results of studies with small sample sizes differed systematically from the results of studies with larger sample sizes (27). To determine the statistical significance of funnel plot asymmetry, Egger's test was then performed; this test is a regression analysis of effect size on its standard error weighted by inverse variance (28). Specifically, diagonal lines indicate that 95% of studies are expected to lie in the triangular region reflecting the absence of both biases and heterogeneity. Evidence of publication bias consisted of either an asymmetrical funnel plot or a *P*-value of less than 0.05 for the Egger's test (29).

## Results

### *Systematic Review of the Studies*

### *Study Characteristics*

Characteristics of the 24 studies are summarized in Appendix 2. All the studies had been published since 2017, and they included a total of 27,982 subjects. Nine studies were conducted in community settings (20-22, 30-35) and 13 studies were in outpatient settings (36-48), with only two studies being performed in inpatient settings (19, 45). Five studies employed a longitudinal design (37, 44, 45, 48, 49), and the other 19 studies used a cross-sectional design. Study populations included undergraduate college students (35), adults (21, 22), family caregivers (39, 40), and people with diseases (19, 20, 30-34, 37, 38, 41-49).

### *Approaches to Identify Symptom Cluster*

Four studies used variable-centered approaches with one of two statistical methods, either network analysis (22, 32) or principal component analysis (43, 44). The remaining 20 studies used person-centered approaches; the most frequently used was latent profile analysis (19, 20, 30, 33, 34, 36, 41, 45, 47-49), followed by cluster analysis (21, 39, 40, 42), latent class analysis (31, 35, 38), and k-means clustering (37, 46).

### *PROMIS Measures*

Characteristics of PROMIS measures used in the 24 studies are summarized in Appendix 3. Various versions of PROMIS measures addressing 20 domains were used for symptom cluster research (Fig. 2). Anxiety and depression were the most frequently used domains, followed by fatigue, sleep disturbance, pain interference, physical function, and satisfaction with social roles and activities. PROMIS computerized adaptive tests (CAT) and short forms were the most frequently used versions of the measures.

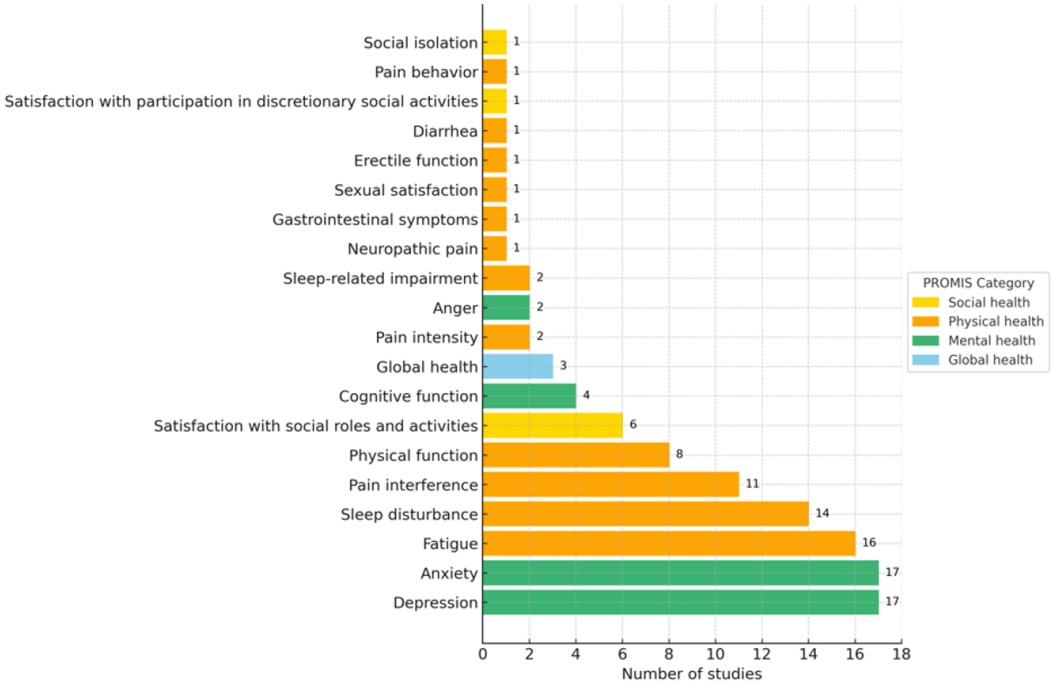


**Fig. 2: Categories and domains of PROMIS measures used in the 24 included studies.** *Note.* Twenty PROMIS domains were grouped into four categories according to the PROMIS classification. PROMIS=Patient-Reported Outcomes Measurement Information System

### *Quality Appraisal*

The findings of the study quality appraisal are summarized in Appendix 4. As to the assessment of risk of bias, the flow and timing domain showed the highest risk of bias across the 24 studies, followed by the patient selection, index test, and reference standard domains. Specifically, 17 studies showed low risk of bias for the matter of patient selection, and six showed high risk of bias for avoiding inappropriate exclusions. With regard to applicability concerns, seven of the 24 studies showed high applicability concern because the research questions of the current review did not match with those studies' participants. All 24 studies showed low applicability concern with respect to the index test and reference standard domains.

### *Reliability Generalization Meta-analysis*

Only four studies that reported internal consistency reliability values were included for meta-analysis, and these studies are summarized in Table 1. Among the 4,131 adult participants in these four studies, 3,981 were community-dwelling adults, and 1,650 were individuals with diseases or health conditions.

**Table 1:** Summary of studies included in the meta-analysis (N=4)

| Author | Year | Sample | Cronbach's alpha value for PROMIS domains | | | | |
|---|---|---|---|---|---|---|---|
| | | | Anxiety | Depression | Fatigue | Pain | Sleep disturbance |
| Breazeale et al. | 2022 | 150 | 0.88 | 0.78 | | 0.69 | 0.88 |
| Lee et al. | 2020 | 1,500 | | 0.97 | 0.96 | 0.98 | 0.95 |
| Starcevic et al. | 2019 | 751 | 0.94 | | | | |
| Shensa et al. | 2018 | 1,730 | 0.90 | 0.93 | | | |

*Note.* PROMIS=Patient-Reported Outcomes Measurement Information System

### *Internal Consistency Reliability*

Cronbach's alpha values reported by four studies were used to calculate the overall mean internal consistency reliability for all the PROMIS domains evaluated—anxiety, depression, fatigue, pain, and sleep disturbance. Because some of the studies reported more than one Cronbach's alpha value for anxiety, depression, pain, and sleep disturbance, these domains were subjected to individual meta-analysis. In contrast, a Cronbach's alpha value for the fatigue domain was reported in only one study (20); therefore, it was excluded from the individual meta-analysis but was included in the calculation of the overall mean internal consistency reliability score.

The reliability generalization meta-analysis showed significant heterogeneity across the four studies included. Except for the sleep disturbance domain, Q-statistics were significant ($P$<.05), indicating significantly greater variability across studies than could be explained by sampling error alone. The Q-statistic for all five PROMIS domains was 106.59, and individual Q-statistics ranged from 3.09 for sleep disturbance to 25.33 for depression. In other words, the depression domain showed considerable variability, whereas the sleep disturbance domain was relatively consistent. In addition, the overall $I^2$ value for all the PROMIS domains was 97.65%, also suggesting substantial heterogeneity across studies.

### *All PROMIS domains*

Five PROMIS domains (anxiety, depression, pain, sleep disturbance, and fatigue) were included in calculation of overall internal consistency reliability. Based on the random effects model, the overall mean score was 0.91 (95% confidence interval [CI] [0.87 to 0.95], $I^2$=97.65%, $P$ <.001), indicating good internal consistency reliability and substantial heterogeneity (Fig. 3).

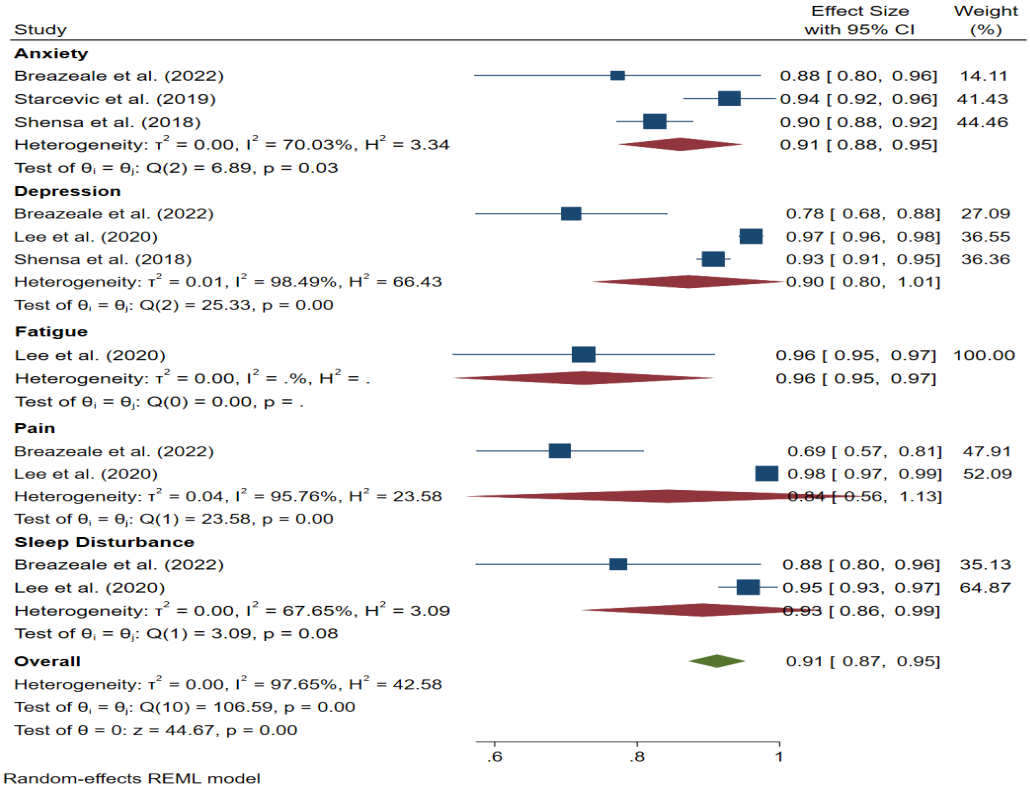


**Fig. 3: Forest plots for PROMIS anxiety, depression, fatigue, pain, sleep disturbance, and overall.** *Note.* CI=confidence interval, REML = Restricted Maximum Likelihood

Cronbach's alpha values from three studies were included in calculation of overall internal consistency reliability for PROMIS anxiety (19, 21, 22). The mean score was 0.91 (95% CI [0.88 to 0.95], $I^2$=70.03%, *P*=.03), indicating good internal consistency reliability and high heterogeneity.

Three studies' Cronbach's alpha values were included in calculation of overall internal consistency reliability for PROMIS depression (19-21). The mean score was 0.90 (95% CI [0.80 to 1.01], $I^2$=98.49%, *P*<.001), indicating good internal consistency reliability and substantial heterogeneity.

Only one study reported a Cronbach's alpha value for PROMIS fatigue (20), no between-study meta-analysis could be conducted. However, this single study reported high internal consistency reliability (α = 0.96).

Cronbach's alpha values from two studies were included in calculation of overall internal consistency reliability for PROMIS pain (19, 20). The mean score was 0.84 (95% CI [0.56 to 1.13], $I^2$=95.76%, *P*<.001), indicating good internal consistency reliability and substantial heterogeneity.

Two studies' Cronbach's alpha values were employed in calculation of overall internal consistency reliability for PROMIS sleep disturbance (19, 20). The mean score was 0.93 (95% CI [0.86 to 0.99]) while the Q-test for heterogeneity was not statistically significant ($I^2$=67.65%, *P*=.08). These findings suggest good internal consistency reliability. Although some heterogeneity was observed across studies, it was not statistically significant.

### *Publication Bias*

The funnel plots of the PROMIS domains indicate asymmetric distribution of the data, suggesting publication bias (Appendix 5). The statistical significance of the Egger's test for all PROMIS domains (*P*<.001) also indicated publication bias. This is illustrated by the lack of studies within the bottom right-hand corner of the funnel plots.

## Discussion

This systematic review and meta-analysis provide the first synthesis of the internal consistency reliability of PROMIS measures applied in adult symptom cluster research. Of the 24 studies included in our review, none reported validity evidence specific to this context, and only four reported internal consistency reliability values. In the meta-analysis, PROMIS anxiety, depression, pain, fatigue, and sleep disturbance demonstrated good to excellent internal consistency on average (overall α=0.91). However, we identified substantial between-study heterogeneity—especially for depression and pain—indicating that reliability varies meaningfully by context (e.g., study population, which PROMIS measures/forms were used, and how they were administered). By bringing together findings across diverse populations and clinical settings, it offers important insights into current applications of PROMIS, highlights domains with stronger psychometric support, and identifies key gaps for future investigations. These strengths enhance the contribution of our study to symptom science and biobehavioral nursing research.

Our meta-analysis revealed good internal consistency reliability for five commonly used PROMIS domains anxiety, depression, fatigue, pain, and sleep disturbance. Outside symptom cluster research, PROMIS measures have demonstrated validity in general clinical and population health research (50-52). More specifically, PROMIS fatigue and PROMIS depression have demonstrated both reliability and validity across a variety of clinical and nonclinical populations (50, 51). Additionally, PROMIS physical function has been shown to be a sensitive and valid tool for detecting changes in physical function across diverse disease populations (53). Furthermore, a recent systematic review evaluating the reliability and validity of PROMIS-29 for assessing health-related quality of life in adults in Switzerland concluded that although this tool is potentially recommendable, further validation is required

due to limited evidence of its psychometric properties (54). Among the studies included in our review, none assessed other types of reliability such as test-retest, interrater, or intrarater reliability, limiting our ability to draw broader conclusions about overall measurement stability. According to classical measurement theory, reliability can also be influenced by participant characteristics and random error, which may further impact measurement consistency (55). Moreover, none of the studies in our review reported construct or criterion validity specific to symptom clusters. Nevertheless, PROMIS measures were used to assess symptoms across a broad range of chronic conditions (e.g., cardiovascular disease, cancer, liver disease, cerebrovascular disease, and kidney disease), as well as among caregivers and general adult populations. Therefore, future studies should examine multiple forms of reliability and explicitly assess validity when applying PROMIS measures to symptom cluster research.

PROMIS offers several administration formats—such as fixed-length short forms, computerized adaptive tests (CAT), profiles, and scales—that may improve accessibility and precision in various populations. In our review, participants completed PROMIS tools either through paper-based short forms or CATs. These formats differ in their psychometric strengths. CATs leverage item response theory to tailor items to each respondent, increasing measurement precision and reducing response burden (56, 57). In contrast, fixed short forms sum item responses without adaptive calibration, potentially limiting sensitivity. Future research should compare reliability and validity across PROMIS formats to inform best practices in symptom cluster research.

Across the 24 studies reviewed, 20 different PROMIS domains were used, reflecting the system's breadth and flexibility. Five domains—anxiety, depression, fatigue, pain interference, and sleep disturbance—were most commonly used and may represent core symptom constructs across conditions. However, none of these domains has been formally validated for their use in identifying or characterizing symptom clusters. Additional studies are needed to assess their factor structure, measurement invariance, and responsiveness to interventions when used in cluster-based symptom science.

Methodological inconsistencies were also identified in how symptom clusters were defined. Most studies followed one of two conceptual approaches: variable-centered (e.g., principal component or network analysis) or person-centered (e.g., latent class or latent profile analysis). However, some studies grouped symptoms using cut-off scores or clinical thresholds without statistical justification. While these cut-off–based methods may be pragmatic in clinical settings, they deviate from established clustering methodologies and may limit reproducibility or generalizability (58, 59). Further clarification is needed on whether such approaches align with the conceptual goals of symptom cluster research, which increasingly aims to identify underlying biological or mechanistic pathways. Importantly, statistical clustering methods require adequate sample sizes, which may not always be feasible in preliminary or mechanistic studies.

Our review also highlights gaps in reliability assessment. While internal consistency was reported in four studies, other key forms of reliability—such as test-retest reliability (i.e., measurement invariance over time)—were not evaluated. Internal consistency reflects the degree to which items within a scale are interrelated, typically quantified using Cronbach's alpha (56). However, it does not capture whether symptom cluster assessments remain stable over time or across raters. As the structure of symptom clusters may evolve, especially in response to treatment, test-retest reliability will be essential to evaluate the longitudinal utility of PROMIS domains in this context.

### *Limitations*

Several limitations of this review should be acknowledged. First, the generalizability of the findings may be constrained by the exclusion of grey literature and the restriction to studies published in English across four major databases. Although we employed comprehensive search strategies and manually reviewed reference lists to

minimize omissions, it is possible that relevant studies were missed. Future reviews should incorporate a broader range of studies to minimize selection bias and enhance the generalizability of their findings. Second, only four studies reported internal consistency reliability, and none reported other psychometric properties such as test-retest or interrater reliability, nor any form of validity. As a result, the psychometric synthesis was limited in scope and depth. Third, due to the small number of studies eligible for meta-analysis, we were unable to conduct subgroup or moderator analyses to explore variability in reliability estimates. To address these limitations, future symptom cluster researchers should rigorously evaluate and report the reliability and validity of the PROMIS measures employed, which would allow more robust subgroup and moderator analyses to clarify factors influencing reliability estimates. Finally, signs of publication bias—evidenced by funnel plot asymmetry and a significant Egger's test—suggest that studies with null or lower reliability estimates may be underrepresented. Given the evidence of publication bias, our findings regarding the reliability of PROMIS tools should be interpreted with caution.

## Conclusion

This systematic review and meta-analysis provide the first synthesis of the internal consistency reliability of PROMIS measures used in adult symptom cluster research. Our findings support the reliability of commonly used PROMIS domains in symptom cluster research, specifically anxiety, depression, fatigue, pain, and sleep disturbance, while also highlighting the need for further psychometric evaluation given the limited number of eligible studies and their high heterogeneity. This study not only maps the current use of PROMIS in symptom cluster research, but by identifying domains with stronger internal consistency, it also offers researchers and clinicians practical guidance for application of PROMIS tools to assess patient-reported outcomes and design evidence-based interventions.

## Journalism Ethics considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

## Acknowledgements

We are thankful to Mr. Jon Mann of the University of Illinois Chicago, who supplied editorial assistance during the paper's preparation. We are also grateful to Ms. Rebecca Raszewski, Health Sciences Librarian of the University of Illinois Chicago, who offered valuable guidance for the literature search.

## Funding Statement

This research was supported by the Chung-Ang University Research Grants in 2023.

## Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Availability of supplementary data

All supplementary data, are present in journal website. Besides, they are accessible via sending email to the corresponding author based on reasonable application.

## References

1. Miaskowski C, Dodd M, Lee K (2004). Symptom clusters: the new frontier in symptom management research. *J Natl Cancer Inst Monogr,* 32:17-21.
2. Barsevick AM (2007). The elusive concept of the symptom cluster. *Oncol Nurs Forum,* 34 (5):971-980.

3. Miaskowski C (2016). Future directions in symptom cluster research. *Semin Oncol Nurs*, 32 (4):405-415.
4. Badger TA, Heitkemper M, Lee KA, et al (2014). An experience with the Patient-Reported Outcomes Measurement Information System: pros and cons and unanswered questions. *Nurs Outlook,* 62 (5):332-338.
5. Alonso J, Bartlett SJ, Rose M, et al (2013). The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. *Health Qual Life Outcomes,* 11:210.
6. Gruber-Baldini AL, Velozo C, Romero S, et al (2017). Validation of the PROMIS® measures of self-efficacy for managing chronic conditions. *Qual Life Res,* 26 (7):1915-1924.
7. Rothrock NE, Amtmann D, Cook KF (2020). Development and validation of an interpretive guide for PROMIS scores. *J Patient Rep Outcomes*, 4 (1):16.
8. Cella D, Riley W, Stone A, et al (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*, 63 (11):1179-1194.
9. Yount SE, Cella D, Blozis S (2019). PROMIS®: Standardizing the patient voice in health psychology research and practice. *Health Psychol,* 38 (5):343-346.
10. Terwee CB, Zuidgeest M, Vonkeman HE, et al (2021). Common patient-reported outcomes across ICHOM Standard Sets: the potential contribution of PROMIS®. *BMC Med Inform Decis Mak,* 21 (1):259.
11. Kroenke K, Haggstrom D, Monahan P, et al. (2018). *Does Sharing Patient-Reported Outcomes with Doctors Improve Patient Symptoms?*. Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/5.2018.ME.140312043
12. Corwin EJ, Moore SM, Plotsky A, et al (2017). Feasibility of combining common data elements across studies to test a hypothesis. *J Nurs Scholarsh,* 49 (3):249-258.
13. Jones FJS, Ezzeddine FL, Herman ST, et al (2020). A feasibility assessment of functioning and quality-of-life patient-reported outcome measures in adult epilepsy clinics: A systematic review. *Epilepsy Behav*, 102:106704.
14. Kalantar-Zadeh K, Lockwood MB, Rhee CM, et al (2022). Patient-centred approaches for the management of unpleasant symptoms in kidney disease. *Nat Rev Nephrol,* 18(3):185-198.
15. Evans JP, Smith A, Gibbons C, et al (2018). The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK. *Patient Relat Outcome Meas*, 9:345-352.
16. Cooper H (2015). *Research synthesis and meta-analysis: A step-by-step approach*. 5th ed. Sage publications, Inc.
17. Elsman EBM, Butcher NJ, Mokkink LB, et al (2022). Study protocol for developing, piloting and disseminating the PRISMA-COSMIN guideline: a new reporting guideline for systematic reviews of outcome measurement instruments. *Syst Rev*, 11 (1):121.
18. Prinsen CAC, Mokkink LB, Bouter LM, et al (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res,* 27 (5):1147-1157.
19. Breazeale S, Conley S, Jeon S, et al (2022). Symptom cluster profiles following traumatic orthopaedic injuries. *Injury,* 53 (7):2524-2532.
20. Lee L, Ross A, Griffith K, et al (2020). Symptom clusters in breast cancer survivors: A latent class profile analysis. *Oncol Nurs Forum*, 47 (1):89-100.
21. Shensa A, Sidani JE, Dew MA, et al (2018). Social media use and depression and anxiety symptoms: A cluster analysis. *Am J Health Behav*, 42 (2):116-128.
22. Starcevic V, Baggio S, Berle D, et al (2019). Cyberchondria and its relationships with related constructs: A network analysis. *Psychiatr Q,* 90 (3):491-505.
23. de Graaff AM, Cuijpers P, Leeflang M, et al (2021). A systematic review and meta-analysis of diagnostic test accuracy studies of self-report screening instruments for common mental disorders in Arabic-speaking adults. *Glob Ment Health (Camb),* 8:e43.
24. Barrientos-Trigo S, Gil-García E, Romero-Sánchez JM, et al (2019). Evaluation of psychometric properties of instruments measuring nursing-sensitive outcomes: a systematic review. *Int Nurs Rev*, 66(2):209-223.

25. Whiting PF, Rutjes AW, Westwood ME, et al (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med,* 155 (8):529-536.
26. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, et al (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods,* 11 (2):193-206.
27. Pustejovsky JE, Rodgers MA (2019). Testing for funnel plot asymmetry of standardized mean differences. *Res Synth Methods,* 10 (1):57-71.
28. Egger M, Davey SG, Schneider M, et al (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ,* 315 (7109):629-634.
29. Sterne JA, Egger M (2001). Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol,* 54 (10):1046-1055.
30. Conley S, Jeon S, Proctor DD, et al (2018). Longitudinal changes in symptom cluster membership in inflammatory bowel disease. *J Nurs Scholarsh,* 50 (5):473-481.
31. Conley S, Proctor DD, Jeon S, et al (2017). Symptom clusters in adults with inflammatory bowel disease. *Res Nurs Health,* 40 (5):424-434.
32. Henneghan A, Wright ML, Bourne G, et al (2021). A cross-sectional exploration of cytokine-symptom networks in breast cancer survivors using network analysis. *Can J Nurs Res,* 53 (3):303-315.
33. Reeve BB, Tan X, Chen RC, et al (2018). Symptom and function profiles of men with localized prostate cancer. *Cancer,* 124 (13):2832-2840.
34. Stone AL, Holley AL, Dieckmann NF, et al (2019). Use of the PROMIS-29® to identify subgroups of mothers with chronic pain. *Health Psychol,* 38 (5):422-430.
35. Vivier H, Ross EJ, Cassisi JE (2020). Classification of gastrointestinal symptom patterns in young adults. *BMC Gastroenterol,* 20 (1):326.
36. Katzan IL, Schuster A, Bain M, et al (2019). Clinical symptom profiles after mild-moderate stroke. *J Am Heart Assoc*, 8 (11):e012421.
37. Kent ML, Giordano NA, Rojas W, et al (2022). Multidimensional perioperative recovery trajectories in a mixed surgical cohort: A longitudinal cluster analysis utilizing National Institutes of Health Patient-Reported Outcome Measurement Information System measures. *Anesth Analg*, 134 (2):279-290.
38. Lapin B, Davin S, Stilphen M, et al (2022). Stratification of spine patients based on self-reported clinical symptom classes. *Spine J*, 22 (7):1131-1138.
39. Lee LJ, Son H, Wallen GR, et al (2023). Symptom clusters in family caregivers of hematopoietic stem cell transplantation recipients: Loneliness as a risk factor. *Transplant Cell Ther,* 29 (1):50.e1-50.e8.
40. Lee LJ, Wehrlen L, Wallen GR, et al (2021). Symptom clusters and influencing factors in family caregivers of individuals with cancer. *Cancer Nurs,* 44 (6):E547-E555.
41. Li H, Lockwood MB, Schlaeger JM, et al (2023). Tryptophan and kynurenine pathway metabolites and psychoneurological symptoms among breast cancer survivors. *Pain Manag Nurs*, 24 (1):52-59.
42. Matura LA, Fargo JD, Boyle K, et al (2022). Symptom phenotypes in pulmonary arterial hypertension: The PAH "symptome". *Pulm Circ,* 12(3):e12135.
43. McDowell L, Casswell G, Bressel M, et al (2021). Symptom burden, quality of life, functioning and emotional distress in survivors of human papillomavirus associated oropharyngeal cancer: An Australian cohort. *Oral Oncol*, 122:105560.
44. Patidar KR, Thacker LR, Wade JB, et al (2017). Symptom domain groups of the Patient-Reported Outcomes Measurement Information System tools independently predict hospitalizations and re-hospitalizations in cirrhosis. *Dig Dis Sci*, 62 (5):1173-1179.
45. Lapin B, Li Y, Englund K, et al (2024). Health-related quality of life for patients with post-acute COVID-19 syndrome: Identification of symptom clusters and predictors of long-term outcomes. *J Gen Intern Med*, 39 (8):1301-1309.
46. Goto T, Saligan LN, Juneau P, et al (2024). Characterization of cancer survivors clustered by subjective and objective cognitive function scores. *Cancer Med,* 13 (12):e7255.
47. Han CJ, Saligan L, Crouch A, et al (2023). Latent class symptom profiles of colorectal cancer survivors with cancer-related

cognitive impairment. *Support Care Cancer,* 31 (10):559.
48. Troost JP, Gipson DS, Carlozzi NE, et al (2019). Using PROMIS® to create clinically meaningful profiles of nephrotic syndrome patients. *Health Psychol,* 38 (5):410-421.
49. Rimmele DL, Schrage T, Lebherz L, et al (2021). Profiles of patients' self-reported health after acute stroke. *Neurol Res Pract,* 3 (1):43.
50. Ameringer S, Elswick RK Jr, Menzies V, et al (2016). Psychometric evaluation of the Patient-Reported Outcomes Measurement Information System fatigue-short form across diverse populations. *Nurs Res,* 65 (4):279-289.
51. Stone AA, Broderick JE, Junghaenel DU, et al (2016). PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity. *J Clin Epidemiol*, 74:194-206.
52. Vilagut G, Forero CG, Adroher ND, et al (2015). Testing the PROMIS® Depression measures for monitoring depression in a clinical sample outside the US. *J Psychiatr Res,* 68:140-150.
53. Schalet BD, Rothrock NE, Hays RD, et al (2016). Validity of PROMIS physical function measures in diverse clinical samples. *J Clin Epidemiol*, 73:112-118.
54. Elsener S, Steinemann N, Wertli MM, et al (2025). A COSMIN systematic review of generic patient-reported outcome measures in Switzerland. *Qual Life Res*, 34:1869-1884.
55. Waltz CF, Strickland O, Lenz ER, et al (2016). *Measurement in Nursing and Health Research.* 5th ed. Springer Publishing Company, New York.
56. DeVellis RF (2016). *Scale development: Theory and applications.* 4th ed. Sage Publications, Inc.
57. Yu L, Buysse DJ, Germain A, et al (2011). Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. *Behav Sleep Med*, 10 (1):6-24.
58. Davis LL, Kroenke K, Monahan P, et al (2016). The SPADE symptom cluster in primary care patients with chronic pain. *Clin J Pain*, 32 (5):388-393.
59. Feng LR, Dickinson K, Kline N, et al (2016). Different phenotyping approaches lead to dissimilar biologic profiles in men with chronic fatigue after radiation therapy. *J Pain Symptom Manage,* 52 (6):832-840.