



REVIEW

A systematic review of patient-reported outcome measures used to assess sleep in postpartum women using Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines

P. Sultan^{1,◉}, K. Ando^{1,*}, E. Sultan¹, J. Hawkins², L. Blake^{3,◉}, F. Barwick⁴, M. Kawai⁴ and B. Carvalho¹

¹Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA, ²Stanford University School of Medicine, Stanford, CA, USA, ³UAMS Medical Library, UAMS, Little Rock, AR, USA and ⁴Division of Sleep Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University, School of Medicine, Stanford, CA, USA

*Corresponding author. Pervez Sultan, 300 Pasteur Drive, Stanford University School of Medicine, CA 94305, USA. Email: psultan@stanford.edu.

Abstract

Study Objectives: We performed a systematic review to identify the best patient-reported outcome measure (PROM) of postpartum sleep in women.

Methods: We searched four databases for validated PROMs used to assess postpartum sleep. Studies were considered if they evaluated at least one psychometric measurement property of a PROM. An overall performance rating was assigned for each psychometric measurement property of each PROM based upon COSMIN criteria. A modified GRADE approach was used to assess the level of evidence and recommendations were then made for each PROM.

Results: We identified 15 validation studies of eight PROMs, in 9,070 postpartum women. An adequate number of sleep domains was assessed by five PROMs: Bergen Insomnia Scale (BIS), Pittsburgh Sleep Quality Index (PSQI), General Sleep Disturbance Scale (GSDS), Athens Insomnia Scale (AIS), and the Sleep Symptom Checklist (SSC). BIS and GSDS were the only PROMs to demonstrate adequate content validity and at least a low level of evidence of sufficient internal consistency, resulting in Class A recommendations. The BIS was the only PROM, which is easily accessible and free to use for noncommercial research, that achieved a Class A recommendation.

Conclusion: The BIS is the best currently available PROM of postpartum sleep. However, this PROM fails to assess several important domains such as sleep duration (and efficiency), chronotype, sleep-disordered breathing and medication usage. Future studies should focus on evaluating the psychometric measurement properties of BIS in the North American setting and in different cultural groups, or to develop a more specific PROM of postpartum sleep.

Key words: postpartum; sleep; patient-reported outcome measure; PROM; COSMIN

Submitted: 1 February, 2021; Revised: 2 April, 2021

© Sleep Research Society 2021. Published by Oxford University Press on behalf of the Sleep Research Society. All rights reserved. For permissions, please email: journals.permissions@oup.com

Introduction

An estimated 140 million women globally recover from child-birth per year [1]. Sleep disturbance is commonly experienced in mothers with a healthy newborn baby [2–5], and sleep is known to be a key domain of postpartum recovery [6]. Women frequently report tiredness and lack of sleep beyond 12 months postpartum [7, 8], with approximately 7% of women requiring some form of treatment or specialist referral in the first 18 months postpartum [8]. Sleep loss has been implicated in diminished well-being for new mothers [9], and is likely to influence other key domains of recovery including psychosocial distress (e.g. depression and psychological morbidity) [10–13], psychosocial support (e.g. marital relations, family dysfunction) [14, 15], maternal–neonatal bonding [16], and fatigue [17]. Maternal suicide remains the second commonest cause of postpartum mortality within the United States [18], and therefore interventions aimed at improving sleep may also help to reduce the incidence and severity of postpartum depression, and associated morbidity [10–12].

While postpartum sleep disturbance is common, there remains no consensus regarding which patient-reported outcome measure (PROM; a structured questionnaire allowing patients to report their health status) should be used to evaluate this complex multidimensional construct. A recent scoping review of postpartum recovery measures identified 8 validated PROMs, which have been used to evaluate postpartum sleep [6]. However, the scoping review did not evaluate the quality of studies or the psychometric measurement properties of the 8 PROMs identified. Systematic reviews can help summarize psychometric measurement properties of PROMs and provide evidence-based recommendations regarding which PROM to select for a given purpose [19, 20]. Reviews can also facilitate future study design, highlight knowledge gaps pertaining to psychometric measurement properties of existing PROMs, and identify the need for new PROMs.

In this systematic review, we utilize Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines [21] to evaluate the psychometric measurement properties of existing PROMs used to assess sleep in postpartum women and make recommendations regarding which is the best available measure.

Methods

This study follows the COSMIN guidelines for performing systematic reviews of PROMs [21], and was registered with PROSPERO (CRD42020223767). As sleep domains in the postpartum period are undefined, postpartum recovery and sleep experts within the author group (P.S., B.C., F.B., M.K.) developed a list of sleep domains that were deemed most relevant to sleep in postpartum women in the outpatient setting. The list of sleep domains was based upon literature review, professional society recommendations (The American Academy of Sleep Medicine, World Sleep Society, National Sleep Foundation and European Sleep Research Society Guidelines) and clinical reasoning [22–30]. The writing committee agreed upon eleven domains of sleep (Supplementary Table 1) including: sleep quality, excessive daytime sleepiness, sleep duration or efficiency, sleep latency, sleep arousal, sleep latency after arousal, chronotype, sleep disordered breathing, subjective restorative feeling, need for hypnotics and sleep interference with daytime functioning.

Search strategy

A medical librarian (L.B.) performed an electronic search of the literature using the following databases: PubMed, Web of Science, CINAHL and Embase. The search was performed with no date limiters on July 28, 2020 to identify all postpartum studies utilizing any of the following 8 validated multi-item (>1 question) PROMs: Bergen Insomnia Scale (BIS), Pittsburgh Sleep Quality Index (Pittsburgh SQI), General Sleep Disturbance Scale (GSDS), Postpartum Sleep Quality Scale (PSQS), Insomnia Severity Index (ISI), Epworth Sleepiness Scale (ESS), Athens Insomnia Scale (AIS) and the Sleep Symptom Checklist (SSC). These PROMs have been used to assess sleep in the postpartum period and were identified in a previously published scoping review [6]. Supplementary File 1 summarizes the search strategy used to identify postpartum studies that evaluated sleep following all delivery modes utilizing these PROMs.

Inclusion and exclusion criteria

We included studies assessing one or more psychometric measurement property of a given PROM of sleep in the postpartum setting. We included randomized controlled trials as well as prospective and retrospective studies. At least one of the following 8 psychometric measurement properties described by the COSMIN group required evaluation in a postpartum study in order for it to be termed a validation study: [21] structural validity (whether scores of a PROM adequately reflect the dimensions of postpartum sleep), internal consistency (interrelatedness among individual PROM items), cross cultural validity (whether performance of items on a translated or culturally adapted PROM adequately reflect performance of the original version) / measurement invariance (whether item responses by women from different groups are similar), reliability (extent to which scores in patients who have not changed are the same when repeated measures are performed), measurement error (systemic and random error of individual scores not attributed to true changes in postpartum sleep), criterion validity (whether a score adequately reflects a “gold standard” measure), hypothesis testing (whether scores are consistent with stated study hypotheses or aims, e.g. difference between delivery modes) and/or responsiveness (change in score over time ≥ 2 postpartum time points, i.e. ability to detect a change).

Scores where patients were asked to self-report sleep with an isolated numerical value (either using a verbal reporting or visual analog scale, e.g. 0–10 or 0–100) were excluded as this involves only one question or item to evaluate sleep. We excluded studies that utilized the PROM as an isolated outcome measure and studies that failed to assess any of the above 8 psychometric measurement properties. We also excluded studies that failed to utilize or report PROMs in their entirety (e.g. studies that used a proportion or unvalidated short form of a PROM), studies not published in English, and studies published in the form of theses, letters, editorials and abstracts.

Data extraction

Duplicate publications from different databases were removed, and remaining articles were entered into the Rayyan online reviewing system for evaluation. Four authors (P.S., K.A., E.S., and

J.H.) determined whether studies met the inclusion criteria for this review. A standardized database was used to extract data from the included studies.

We analyzed data from identified validation studies using COSMIN guidelines. Analysis involved PROM assessment in seven steps outlined below [21]:

(1) Content validity.

Content validity was assessed by reviewing the number of sleep domains evaluated by each PROM. A PROM was considered to have adequate content validity if it evaluated at least 6 out of the 11 proposed domains in the postpartum period (i.e., assessment of >50% of the proposed domains).

(2) Feasibility.

Feasibility assessment of individual PROMs was performed by evaluating accessibility or availability, cost of use for noncommercial research, response rates among studies, completion rates of returned PROMs (missingness), and time taken to complete the PROM within studies.

(3) Risk of bias.

Risk of bias was assessed for each validation study in relation to the eight psychometric measurement properties outlined by the COSMIN guidelines. The methodology used in each study to assess a measurement property was graded as: “very good,” “adequate,” “doubtful,” “inadequate” or “not applicable.” If a study assessed multiple measurement properties, an overall risk of bias rating related to the methodology utilized was assigned using a “worst score counts” principle [31]. For example, if the methodology used to assess structural validity in a study was deemed to be “adequate” and the assessment of internal consistency was “very good” in the same study, an overall rating of “adequate” would be assigned for the methodology of that particular study.

(4) Assessment of psychometric measurement properties of PROMs from individual studies.

Psychometric measurement properties of PROMs were assessed in individual studies. The psychometric measurement property performance rating strategy for individual studies is provided in [Supplementary Table 2](#). Performance was determined from the results of individual studies, which provided a rating of: sufficient (+), insufficient (-), inconsistent (+/-) or indeterminate (?).

(5) Overall performance rating of psychometric measurement properties of PROMs from all included studies.

Overall performance ratings were provided for all 8 psychometric measurement properties for each PROM. Ratings for overall psychometric measurement properties were: sufficient (+), insufficient (-), inconsistent (+/-) or indeterminate (?). If only 1 study assessed a psychometric measurement property, then the overall performance (quality) assessment was based on ratings from Step 4. For example, if one study evaluating internal consistency of a single PROM reported results consistent with a sufficient (+) rating, then the overall performance rating for this PROM would be sufficient (+). If >1 study assessed the same psychometric measurement property of a PROM, individual performance ratings from Step 4 for each of the included studies were pooled in order to provide an overall quality assessment of the PROM. When pooling results from multiple studies, a PROM was deemed to have a

sufficient overall performance rating (+) if >50% of the individual studies were graded as sufficient (i.e. if the majority of studies were in accordance with this finding). If no studies assessed a specific measurement property, then this was reported as indeterminate (?). For example, if five studies assessed internal consistency of a particular PROM and were rated as sufficient (+) in three studies, insufficient (-) in one study and indeterminate (?) in one study then an overall performance rating of sufficient (+) was awarded, as this was the rating for the majority of studies.

(6) Level of evidence.

A level of evidence was assigned for the psychometric measurement properties of each PROM using a modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach for systematic reviews of clinical trials. GRADE level of evidence was rated as: “high,” “moderate,” “low” or “very low.” [32] As per the GRADE approach, level of evidence was determined by assessment of included studies for: (1) risk of bias (downgraded by one, two, or three levels if serious, very serious or extremely serious risk of bias, respectively); (2) inconsistency (downgraded by one or two levels if serious or very serious inconsistency, respectively); (3) imprecision (downgraded by one or two levels if between 50–100 or <50 study subjects, respectively), and (4) indirectness (downgraded by one or two levels if serious or very serious indirectness, respectively). Studies failing to report numbers of patients screened or eligible, missing data, response rates, or missingness of data from completed PROMs were downgraded for potential risk of bias. Two authors independently graded all studies (P.S. and K.A.), and conflicts were resolved following discussion with a third author (B.C.).

(7) Recommendations.

A table summarizing the findings from the above Steps 1–6 was used to determine the best available PROM(s) of sleep in postpartum women and provide recommendations for future studies. Each PROM was classified with a level of recommendation of A, B, or C as follows: [21, 33] (A) PROMs with adequate content validity (≥ 6 out of the 11 proposed domains) and at least low-quality evidence for sufficient internal consistency. Class A PROMs are recommended for use and results obtained with these PROMs can be trusted; (B) PROMs not meeting criteria for A or C. Class B PROMs have potential to be recommended for use, but require further research to assess their quality; and (C) PROMs with high quality evidence for a psychometric measurement property rated as insufficient. Class C PROMs are not recommended for use.

Results

Figure 1 summarizes the literature search findings. In total, 15 studies evaluating 8 PROMs of sleep in postpartum women were assessed in this review.[11, 12, 16, 34–45] The included measures were used to evaluate sleep in 9,070 postpartum women.

Study summary

Table 1 summarizes the studies grouped by PROMs. Studies were performed in 9 different countries between 2004 and 2020. Patient numbers evaluated in the included studies varied from 21 to 2,386 women. All studies evaluating psychometric

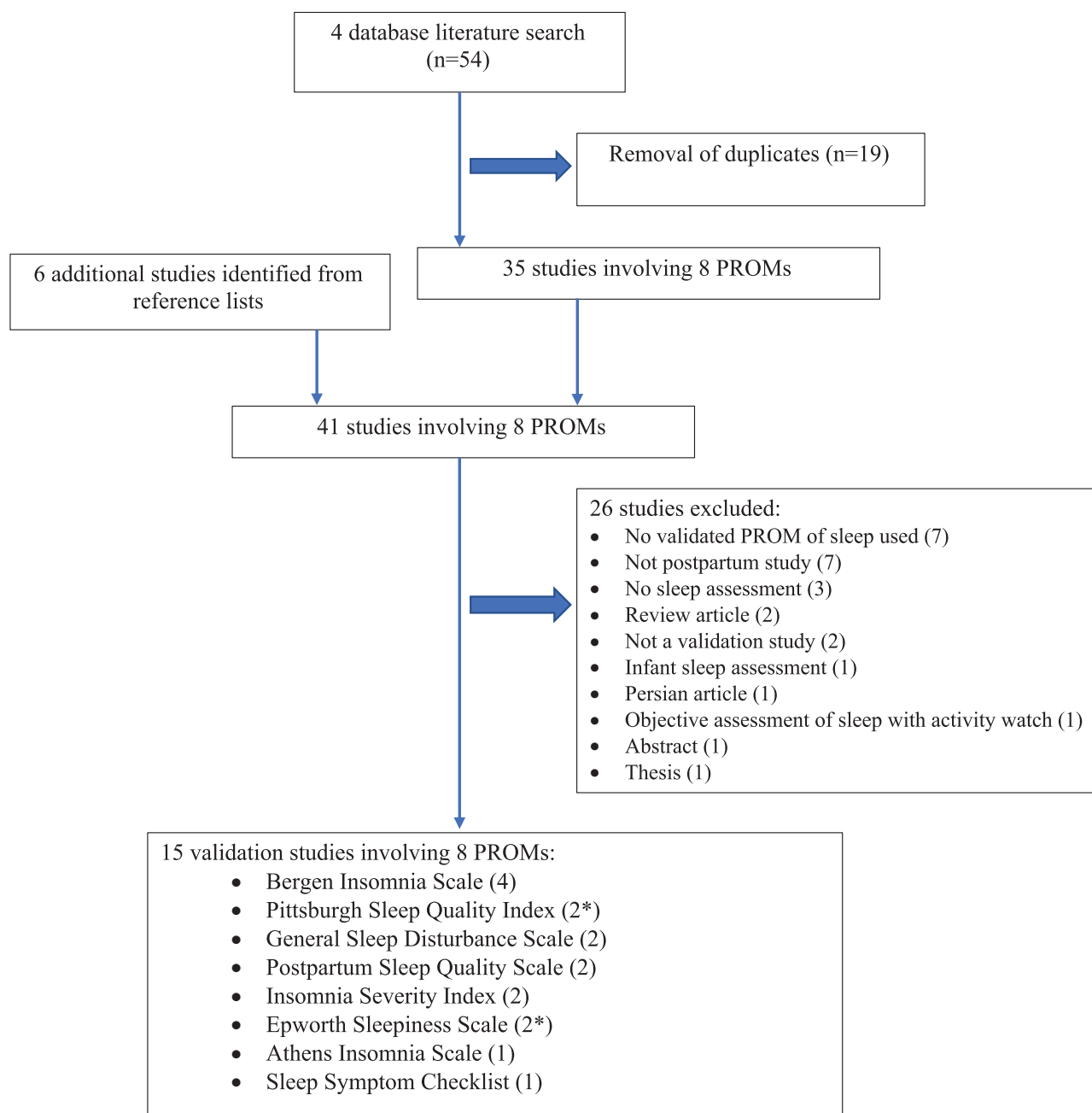


Figure 1. Summary of search. PROMs=patient-reported outcome measures; *one validation study evaluated two included patient-reported outcome measures (Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale).

measurement properties of the included PROMs utilized a prospective study design, and none of the studies were randomized controlled trials. Included studies evaluated sleep from the day following delivery up to 2 years postpartum. Delivery modes of subjects were not reported in the studies evaluating Insomnia Severity Index (ISI), Athens Insomnia Scale (AIS) or Sleep Symptom Checklist (SSC). Of the 8 PROMs evaluated, only one, the Postpartum Sleep Quality Scale (PSQS), was specifically developed for use in postpartum women.[40] Sleep following operative vaginal delivery was not specifically evaluated or reported using any of the included PROMs. [Supplementary Table 2](#) provides a summary of the individual studies including languages of assessed PROMs. Languages

of PROMs used were heterogeneous but most frequently involved English and Norwegian versions (five and four studies respectively).

Content validity

[Table 2](#) provides a summary of the content validity of included PROMs. Seven out of the 8 included PROMs assessed multiple domains (≥ 5 domains) of sleep, with the exception of the Epworth Sleepiness Scale (ESS), which only assesses the domain of excessive daytime sleepiness. Adequate content validity (≥ 6 out of 11 domains) was demonstrated by the

Table 1. Summary of studies utilizing each patient-reported outcome measure

	Obstetric/ postpartum specific	Validation studies done Y/N (n)	Postpartum patients studied (n)	Prospective study Y/N n (n of RCTs)	Language(s) used with instrument	Study year range	Delivery modes VD OVD CD	Postpartum time points assessed
BIS	N	Y (4)	6,493	Y 4 (0)	Norwegian	2014–2020	Y ? Y	8 weeks and 2 years
Pittsburgh SQI	N	Y (2*)	674	Y 2 (0)	Chinese	2004–2019	Y ? Y	Between 13th and 20th day Up to 1 month
GSDS	N	Y (2)	133	Y 2 (0)	English	2007–2009	Y ? Y	Between 3rd and 5th day 1, 2 and 3 months
PSQS	Y	Y (2)	302	Y 2 (0)	Chinese, Turkish	2013–2018	Y ? Y	Up to 2 weeks Postpartum clinic (timing not stated)
ISI	N	Y (2)	158	Y 2 (0)	English, Hebrew	2016–2019	? ? ?	Between 3 and 18 months Timing varied de- pending on when admitted for respite care
ESS	N	Y (2*)	1,167	Y 2 (0)	English	2013	Y ? Y	“Immediate postpartum period” (unclear timing)
Athens IS	N	Y (1)	84	Y 1 (0)	Polish	2017	? ? ?	Between 1 and 2 days
SSC	N	Y (1)	287	Y 1(0)	French or English	2019	? ? ?	7–9 weeks

Validation study defined as a study assessing and reporting one of the eight COSMIN-defined psychometric measurement properties of patient-reported outcome measures; BIS=Bergen insomnia scale; Pittsburgh SQI = Pittsburgh Sleep Quality Index; GSDS = General Sleep Disturbance Scale; PSQS = Postpartum Sleep Quality Scale; ISI=Insomnia Severity Index; ESS - Epworth Sleepiness Scale; Athens IS = Athens Insomnia Scale; SSC = Sleep Symptom Checklist;

Total of 15 validation studies assessing multidimensional sleep measures in the postpartum period: *1 study by Huang et al. (n = 228) provides validation data for both ESS and Pittsburgh SQI measures; ? = delivery mode not stated in methodology or results section of published study

Y = yes and N = no; n = number; RCT = randomized control trial; VD = vaginal delivery; OVD = operative vaginal delivery; CD = cesarean delivery; D = day. Hypothesis considered if a measure statistically compared to another measure or evaluated a difference in health state between two groups, e.g. different delivery mode or peripartum morbidity.

following five PROMs: Bergen Insomnia Scale (BIS), Pittsburgh Sleep Quality Index (PSQI), General Sleep Disturbance Scale (GSDS), Athens Insomnia Scale (AIS) and the Sleep Symptom Checklist (SSC).

Feasibility

Table 2 summarizes PROM availability and associated cost for noncommercial use. The original versions of two included PROMs (GSDS and SSC) are not readily available from links through their associated publications nor through an online search conducted via the Google search engine. PROMs were comprised of between 6 and 21 items and evaluated sleep over the preceding month (4 PROMs: BIS, PSQI, AIS and SSC), preceding 2 weeks (two PROMs: PSQS and ISI), preceding week (one measure: GSDS) or in relation to sleep symptoms “in recent times” (one PROM: ESS). The time taken to complete each PROM was not reported in any of the included studies.

Supplementary Table 3 provides a summary of response rates (range between 32% and 100%), screening for eligibility, and reporting of missing data from completed / returned PROMs. The numbers of women screened or eligible were reported in 5 out of 19 studies [34–37, 42]. All studies reported their response rates

and accounted for withdrawals. Only two of the included studies reported missingness data from completed PROMs [38, 42].

Risk of bias

Risk of bias assessment of methodology used to evaluate psychometric measurement properties of PROMs in individual studies is provided in Supplementary Table 4. All studies were graded as “very good” for methodology utilized to assess psychometric measurement properties using the “worst score counts” principle, except for Yang et al., which was graded as “adequate” (Supplementary Table 3).[40]

Assessment of psychometric measurement properties of PROMs from individual studies and overall ratings

Supplementary Table 4 summarizes the psychometric measurement property performance ratings for each PROM based on results from individual studies and overall ratings based on pooling of results from all studies. No studies assessed cross cultural validity, measurement invariance or measurement error.

Table 2. Summary of content validity and questionnaire format of sleep measures in postpartum women

Measure	Permission for use	No. of items	Time period evaluated	Evaluation of content validity (domains assessed by each measure)				
				Sleep quality	Excessive daytime sleepiness	Sleep duration/efficiency	Sleep latency	Sleep arousal
BIS	Available online	6	Past month	X	X		X	X
PSQI	Available online	18	Past month	X	X	X	X	X
GSDS	Unable to obtain*	21	Past week	X	X	X	X	X
PSQS	Available through author	14	Past 2 weeks	X		X	X	X
ISI	Available online	7	Past 2 weeks	X		X	X	X
ESS	Available online	8	Recent times		X			
Athens IS	Upon request from author	8	Past month	X	X	X	X	X
Sleep SC	Unable to obtain*	21	Past month	X	X	X	X	X

Measure	Sleep latency after arousal	Chronotype	Sleep-disordered breathing	Subjective restorative feeling	Need for hypnotics	Sleep interference with daytime functioning ^y	Total no. of domains assessed
BIS	X			X		X	7
PSQI			X		X	X	8
GSDS					X		6
PSQS						X	5
ISI						X	5
ESS							1
Athens IS						X	6
Sleep SC	X		X	X		X	9

BIS = Bergen insomnia scale; PSQI = Pittsburgh Sleep Quality Index; GSDS = General Sleep Disturbance Scale; PSQS = Postpartum Sleep Quality Scale; ISI = Insomnia Severity Index; ESS = Epworth Sleepiness Scale; Athens IS = Athens Insomnia Scale; SSC = Sleep Symptom Checklist. No monetary charge for above measures available online or through author, when used for noncommercial research purposes.

*Original versions of these measures are not available in their referenced published studies or following online search. Information was extracted based on descriptions provided in the corresponding publications.

^yIncludes daytime physical function, infant care, cognitive function, psychosocial distress/affective function, fatigue, psychosocial wellbeing, sexual function.

Level of evidence

[Supplementary Table 4](#) provides a GRADE level of evidence for the overall performance rating of each psychometric measurement property for each PROM, with justification for downgrading where applicable. Level of evidence was low or very low for all of the psychometric measurement properties of the PROMs assessed. Studies were frequently downgraded for study design (convenience sampling predisposing to selection bias), inadequate response rates (<60%), failure to report numbers of women screened or missing data, or low numbers of study participants resulting in imprecision.

Summary of findings

[Table 3](#) summarizes findings from this review. Of the seven PROMs used to assess multiple domains of postpartum sleep, BIS was the only readily available PROM that demonstrated adequate content validity and at least low-level evidence of internal consistency and therefore received a Class A recommendation. While GSDS also received a Class A recommendation, the original version of this PROM could not be retrieved despite exhaustive efforts. The PSQS, ISI and ESS did not demonstrate sufficient content validity, therefore resulting in Class C

recommendations. Class B recommendations were assigned to the PSQI, AIS, and SSC.

Recommendation

Based on available evidence, the best currently available PROM of postpartum sleep is the BIS. However, it must be noted that this PROM just met the threshold for adequate content validity as it fails to assess the following sleep domains: sleep duration/sleep efficiency, chronotype, sleep-disordered breathing and need for hypnotics. Further studies are needed to evaluate psychometric measurement properties of Class A and Class B PROMs in order to maximize levels of evidence and increase the strength of recommendations.

Discussion

The main finding from this study is that the BIS is the only readily available PROM that received a Class A recommendation using the COSMIN criteria. Therefore, it can be considered the best currently available PROM of postpartum sleep. There was insufficient data regarding internal consistency to support

Table 3. Summary of findings table

Patient-reported outcome measures	Content validity # domains	Structural validity			Internal consistency			Cross-cultural validity/ measurement invariance		
		Methods	Results	LoE	Methods	Results	LoE	Methods	Results	LoE
BIS	6	NA	?	NA	Very good	+	Low	NA	?	NA
Pittsburgh SQI	8	NA	?	NA	Very good	-	Low	NA	?	NA
GSDS	6	NA	?	NA	Very good	+	Low	NA	?	NA
PSQS	5	Adequate	+	Low	Very good	+	Low	NA	?	NA
ISI	5	NA	?	NA	NA	?	NA	NA	?	NA
ESS	1	NA	?	NA	Very good	+	Low	NA	?	NA
Athens IS	6	NA	?	NA	NA	?	NA	NA	?	NA
Sleep SC	9	NA	?	NA	NA	?	NA	NA	?	NA

Patient-reported outcome measures	Reliability			Measurement Error			Hypothesis testing			Responsiveness			Recommendation
	Methods	Results	LoE	Methods	Results	LoE	Methods	Results	LoE	Methods	Results	LoE	
BIS	NA	?	NA	NA	?	NA	Very good	+	Low	Very good	-	Very low	A
Pittsburgh SQI	NA	?	NA	NA	?	NA	Very good	+	Low	NA	?	NA	B
GSDS	NA	?	NA	NA	?	NA	Very good	+	Low	NA	?	NA	A
PSQS	Very good	+/-	Low	NA	?	NA	Very good	+	Low	NA	?	NA	C
ISI	NA	?	NA	NA	?	NA	Very good	+	Low	NA	?	NA	C
ESS	NA	?	NA	NA	?	NA	Very good	+	Low	NA	?	NA	C
Athens IS	NA	?	NA	NA	?	NA	Very good	+	Very low	NA	?	NA	B
Sleep SC	NA	?	NA	NA	?	NA	Very good	+	Low	NA	?	NA	B

BIS = Bergen insomnia scale; Pittsburgh SQI = Pittsburgh Sleep Quality Index; GSDS = General Sleep Disturbance Scale; PSQS = Postpartum Sleep Quality Scale; ISI = Insomnia Severity Index; ESS = Epworth Sleepiness Scale; Athens IS = Athens Insomnia Scale; SSC = Sleep Symptom Checklist.

Methodology reported as either: "Very good," "Adequate," "Doubtful," "Inadequate" or "Not assessed" (NA).

Ratings for overall performance reported as either: sufficient (+), insufficient (-), inconsistent (\pm), or indeterminate (?); LoE = level of evidence using GRADE reported as: "High," "Moderate," "Low," "Very low" or "Not assessed" (NA).

Recommendation: (A) evidence for sufficient content validity (≥ 6 domains assessed) and at least low-quality evidence for sufficient internal consistency (measures can be recommended for use); (B) measures categorized not in A or C (require further evaluation to assess quality prior to recommendation for use); (C) high quality evidence for an insufficient rated measurement property (not recommended for use).

Class A recommendations for the PSQI and AIS, but these PROMs may also warrant further psychometric measurement property evaluation in the postpartum setting.

Clinical relevance

As many as 15% of women experience sleep disturbance up to 6 months following delivery [7], and approximately 55% of women report extreme tiredness up to 18 months postpartum following all delivery modes [8]. Although clearly important, sleep and associated fatigue in the postpartum period are underexplored outcomes. A previous scoping review of 521 postpartum recovery studies identified eight validated PROMs used to assess sleep in the postpartum period [6]. To our knowledge, however, this is the first study that systematically reviews psychometric measurement properties of these validated sleep PROMs in postpartum women, utilizing robust COSMIN methodology. Our analysis provides a detailed summary of these PROMs and their psychometric measurement properties, in addition to providing recommendations regarding the best currently available PROM of sleep in the postpartum period.

This study also highlights gaps in current knowledge relating to the quality assessment of commonly used sleep PROMs in the postpartum population. For example, among the PROMs evaluated, no studies adequately assessed cross cultural validity, measurement invariance or measurement error. Structural validity and responsiveness were only assessed in two PSQS

studies and one BIS study, respectively. The current level of evidence for internal consistency was reported as low for all PROMs across studies evaluating this property, and either low or very low for the remaining psychometric measurement properties evaluated. Among included studies, the BIS was the most frequently used PROM involving the largest patient cohorts, but these studies were all performed using a Norwegian version of the tool. Further studies are needed to evaluate psychometric measurement properties of BIS in the North American setting and in different cultural groups, utilizing an English version and COSMIN endorsed methodology, to provide a higher level of evidence supporting the use of this PROM.

PROMs are being used increasingly to assess postoperative recovery [46]. While sleep disturbance is common in the postpartum period, it is currently an infrequently reported outcome measure in postpartum recovery studies. A scoping review of recovery studies recently reported sleep as an outcome in only 13 out of 521 (2.5%) studies that used PROMs to evaluate postpartum recovery [6]. The small number of available PROMs and the few studies evaluating the psychometric measurement properties of these PROMs are disproportionate to the frequency of childbirth and the estimated incidence of sleep disturbance within the postpartum population.

Furthermore, the relative paucity of PROMs available to assess sleep in postpartum women supports the need for research to profile sleep patterns in the weeks and months following childbirth. Clinicians aiming to optimize postpartum

recovery experience should utilize multidimensional sleep measures such as the BIS to evaluate this important recovery domain. By evaluating sleep and other important recovery domains such as fatigue, pain, physical activity, psychosocial distress and psychosocial support, the relationship between these interrelated recovery domains can be determined and potential therapeutic targets identified in women at risk of experiencing delayed or poor recovery health. For example, the relationship between sleep health and depression has been highlighted in two studies included in this review [11, 12], but rates of referral to sleep specialists and prevalence of sleep disorders diagnosed in this population are unclear and warrant further research.

Research implications

Research is frequently performed with measurement instruments of unknown quality, which can be a waste of resources and potentially unethical [47, 48]. The only validated PROM that was developed specifically for use in the postpartum population was the PSQS, however it demonstrated inadequate content validity when evaluated against the postpartum sleep domains proposed in this review. The PSQS (originally developed in Taiwan) fails to ask questions related to the following postpartum sleep domains: daytime sleepiness, sleep latency after arousal, chronotype, sleep disordered breathing, subjective restorative feeling and need for hypnotic medication. The PSQS would therefore require modification, with the addition of new questions in order to achieve adequate content validity. Any changes made to this PROM would however require further studies to assess its psychometric measurement properties.

The development of new PROMs can be time-consuming and expensive. The use of BIS, identified in this review as the most appropriate and adequately validated PROM of postpartum sleep, will help to maximize efficiency and increase ethicality of future research exploring sleep after childbirth. By standardizing PROMs used in planned postpartum studies, heterogeneity can be reduced, and pooling of data through meta-analysis can be facilitated. The Core Outcome Measures in Effectiveness Trials (COMET) initiative aims to develop and apply consensus-based standardized sets of outcomes, known as core outcome sets, which ultimately reduce heterogeneity among studies. Core outcome sets represent the minimum data that should be measured and reported in clinical trials of a specific condition. Identifying BIS as the best currently available PROM of postpartum sleep can facilitate the development of future core outcome sets for postpartum studies. However, well-designed and adequately powered studies are still needed to definitively evaluate all psychometric measurement properties of BIS and PSQI (which demonstrated better content validity but insufficient internal consistency), and their translated versions in postpartum cohorts of women. This study also highlights the lack of a PROM that is able to comprehensively assess all postpartum sleep domains, which may justify the need for development of a new robustly developed PROM with optimum content validity based on domains proposed in this review. This may be achieved using PROMIS (Patient-Reported Outcomes Measurement Information System) endorsed methodology [49].

Limitations

We acknowledge that the development of author-defined domains of sleep in the postpartum period is itself a subjective process, and the contribution of individual sleep domains may change at different postpartum time points. However, by gaining consensus among specialists in postpartum recovery (P.S. and B.C.) and sleep specialists (F.B. and M.K.), and by using a conservative threshold of ≥ 6 out of the 11 proposed domains to demonstrate adequate content validity, we feel that this methodology allowed us to identify the most appropriate PROMs of sleep for postpartum women. We were unable to assess the psychometric measurement property of criterion validity, as no widely accepted gold standard for measuring sleep in the postpartum period exists. We also acknowledge that the reporting of findings regarding quality of individual study methodology and results using COSMIN guidelines is subjective and open to differences in opinion. We minimized this variability by grading studies independently using two authors and then resolving any conflicts by discussion with a third team member.

Conclusion

In summary, BIS is the best currently available PROM to evaluate sleep in the postpartum period. However, this PROM fails to assess several important sleep domains. Clinicians need to be aware of the strengths and weaknesses of PROMs when choosing the appropriate measure to use in the postpartum setting. Future studies are needed to further evaluate the psychometric measurement properties of BIS and PSQI (PSQI demonstrates better content validity) in the North American setting and in different cultural groups, or to develop a new PROM specifically designed to assess sleep in postpartum women.

Supplementary material

Supplementary material is available at SLEEP online.

Funding

Pervez Sultan is an Arline and Pete Harman Endowed Faculty Scholar of the Stanford Maternal and Child Health Research Institute. Dr. Kawai is supported by the National Institute on Aging of the National Institutes of Health K23AG053465, 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation, and Autism Working Group Award from Mosbacher Family Foundation.

Disclosure Statement

None declared.

References

1. The world counts. <https://www.theworldcounts.com/stories/How-Many-Babies-Are-Born-Each-Day#:~:text=The crude birth rate is 18.9 births per,world population is approximately 7.2 billion and growing>. Published 2020. Accessed August 27, 2020.
2. Brunner DP, et al. Changes in sleep and sleep electroencephalogram during pregnancy. *Sleep*. 1994;17(7):576-582.

3. Gay CL, et al. Sleep patterns and fatigue in new mothers and fathers. *Biol Res Nurs*. 2004;5(4):311–318.
4. Lee KA, et al. REM sleep and mood state in childbearing women: sleepy or weepy? *Sleep*. 2000;23(7):877–885.
5. Shinkoda H, et al. Changes in sleep-wake cycle during the period from late pregnancy to puerperium identified through the wrist actigraph and sleep logs. *Psychiatry Clin Neurosci*. 1999;53(2):133–135.
6. Sultan P, et al. Evaluation of domains of patient-reported outcome measures for recovery after childbirth: a scoping and systematic review. *JAMA Netw Open*. 2020;3(5):e205540.
7. Thompson JF, et al. Prevalence and persistence of health problems after childbirth: associations with parity and method of birth. *Birth*. 2002;29(2):83–94.
8. Glazener CM, et al. Postnatal maternal morbidity: extent, causes, prevention and treatment. *Br J Obstet Gynaecol*. 1995;102(4):282–287.
9. Hiscock H, et al. Infant sleep problems and postnatal depression: a community-based study. *Pediatrics*. 2001;107(6):1317–1322.
10. Bhati S, et al. A systematic review of the relationship between postpartum sleep disturbance and postpartum depression. *J Obstet Gynecol Neonatal Nurs*. 2015;44(3):350–357.
11. Goyal D, et al. Fragmented maternal sleep is more strongly correlated with depressive symptoms than infant temperament at three months postpartum. *Arch Womens Ment Health*. 2009;12(4):229–237.
12. Huang CM, et al. A comparison of sleep and daytime sleepiness in depressed and non-depressed mothers during the early postpartum period. *J Nurs Res*. 2004;12(4):287–296.
13. Bei B, et al. Sleep and mood during pregnancy and the postpartum period. *Sleep Med Clin*. 2015;10(1):25–33.
14. Piteo AM, et al. Postnatal depression mediates the relationship between infant and maternal sleep disruption and family dysfunction. *Early Hum Dev*. 2013;89(2):69–74.
15. Meijer AM, et al. Contribution of infants' sleep and crying to marital relationship of first-time parent couples in the 1st year after childbirth. *J Fam Psychol*. 2007;21(1):49–57.
16. Tikotzky L. Postpartum maternal sleep, maternal depressive symptoms and self-perceived mother-infant emotional relationship. *Behav Sleep Med*. 2016;14(1):5–22.
17. Insana SP, et al. Sleep disturbance and neurobehavioral performance among postpartum women. *Sleep*. 2013;36(1):73–81.
18. Wisner KL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. *JAMA Psychiatry*. 2013;70(5):490–498.
19. Prinsen CA, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials*. 2016;17(1):449.
20. Sharawi N, et al. Evaluation of patient-reported outcome measures of functional recovery following caesarean section: a systematic review using the consensus-based standards for the selection of health measurement instruments (COSMIN) checklist. *Anaesthesia*. 2019;74(11):1439–1455.
21. Mokkink L, et al. COSMIN Methodology for Systematic Reviews of Patient-Reported Outcome Measures (PROMs) User Manual. Amsterdam; 2018. https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf. Accessed April 1, 2021.
22. Watson N, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *J Clin Sleep Med*. 2015;11(6):591–592.
23. Watson N, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *J Clin Sleep Med*. 2015;11(8):931–952.
24. Aurora RN, et al. Quality measure for screening for adult obstructive sleep apnea by primary care physicians. *J Clin Sleep Med*. 2016;12(8):1185–1187.
25. Morgenthaler TI, et al. Measurement of quality to improve care in sleep medicine. *J Clin Sleep Med*. 2015;11(3):279–291.
26. Edinger JD, et al. Quality measures for the care of patients with insomnia. *J Clin Sleep Med*. 2015;11(3):311–334.
27. World Sleep Day. Talk Points. World Sleep Society. <http://worldsleepday.org/usetoolkit/talking-points>. Published 2020. Accessed December 30, 2020.
28. World Sleep Day. Ask the Sleep Experts. World Sleep Society.
29. Hirshkowitz M, et al. National Sleep Foundation's updated sleep duration recommendations: final report. *Sleep Health*. 2015;1(4):233–243.
30. Riemann D, et al. European guideline for the diagnosis and treatment of insomnia. *J Sleep Res*. 2017;26(6):675–700.
31. Terwee CB, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–657.
32. GRADE. GRADE Handbook - Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach.; 2013. [https://med.mahidol.ac.th/ceb/sites/default/files/public/pdf/journal_club/2017/GRADE handbook.pdf](https://med.mahidol.ac.th/ceb/sites/default/files/public/pdf/journal_club/2017/GRADE%20handbook.pdf). Accessed April 1, 2021.
33. Prinsen CAC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
34. Dørheim SK, et al. Can insomnia in pregnancy predict postpartum depression? A longitudinal, population-based study. *PLoS One*. 2014;9(4):e94674.
35. Dørheim SK, et al. Personality and perinatal maternal insomnia: a study across childbirth. *Behav Sleep Med*. 2016;14(1):34–48.
36. Osnes R, et al. Mid-pregnancy insomnia and its association with perinatal depressive symptoms: a prospective cohort study. *Behavioural Sleep Med*. 2020. doi:10.1080/15402002.2020.1743705
37. Sivertsen B, et al. Trajectories of maternal sleep problems before and after childbirth: a longitudinal population-based study. *BMC Pregnancy Childbirth*. 2015;15:129.
38. Wang YY, et al. Screening for postpartum hypomania among Chinese women after childbirth. *Psychiatry Res*. 2019;282:112591.
39. Lee SY, et al. Early postpartum sleep and fatigue for mothers after cesarean delivery compared with vaginal delivery an exploratory study. *J Perinat Neonat Nurs*. 2007;21(2):109–113.
40. Yang CL, et al. Development and validation of the postpartum sleep quality scale. *J Nurs Res*. 2013;21(2):148–154.
41. Boz İ, et al. Testing the psychometric properties of the postpartum sleep quality scale in Turkish women. *J Nurs Res*. 2018;26(6):385–392.
42. Wilson N, et al. More than depression: a multi-dimensional assessment of postpartum distress symptoms before and after a residential early parenting program. *BMC Psychiatry*. 2019;19(1):48.

43. Bourjeily G, et al. Epworth sleepiness scale scores and adverse pregnancy outcomes. *Sleep Breath*. 2013;**17**(4):1179–1186.
44. Drozdowicz-Jastrzębska E, et al. Insomnia, postpartum depression and estradiol in women after delivery. *Metab Brain Dis*. 2017;**32**(6):1913–1918.
45. Solomonova E, et al. Sleep quality is associated with vasopressin methylation in pregnant and postpartum women with a history of psychosocial stress. *Psychoneuroendocrinology*. 2019;**107**:160–168.
46. Myles PS. More than just morbidity and mortality - quality of recovery and long-term functional recovery after surgery. *Anaesthesia*. 2020;**75 Suppl 1**:e143–e150.
47. Ioannidis JP, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;**383**(9912):166–175.
48. Walton M, et al. Clinical outcome assessments: a conceptual foundation – report of the ISPOR Clinical Outcomes Assessment Emerging Good Practices Task Force. *Value Heal*. 2015;**18**(6):741–752.
49. PROMIS. PROMIS® Instrument Development and Validation Scientific Standards Version 2.0. http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. Published 2013. Accessed March 3rd, 2020.