Sleep Medicine 90 (2022) 74–82

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Review Article

A scoping review of behavioral sleep stage classification methods for preterm infants

Anne Bik [a, 1], Chanel Sam [a, 1], Eline R. de Groot [a], Simone S.M. Visser [a], Xiaowan Wang [a], Maria Luisa Tataranno [a], Manon J.N.L. Benders [a], Agnes van den Hoogen [a, b, c], Jeroen Dudink [a, d, *]

[a] Department of Neonatology, University Medical Center Utrecht, Wilhelmina Children's Hospital, Utrecht, the Netherlands
[b] Prinses Maxima Center for Pediatric Oncology, Utrecht, the Netherlands
[c] Utrecht University, Utrecht, the Netherlands
[d] Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 2 September 2021
Received in revised form
1 January 2022
Accepted 5 January 2022
Available online 19 January 2022

Keywords:
Sleep
Preterm infant
Behavior
Observation
Human
Score

ABSTRACT

Background: Sleep is paramount for optimal brain development in infants admitted to the neonatal intensive care unit. Besides (minimally) invasive technical approaches to study sleep in infants, there is currently a large variety of behavioral sleep stage classification methods (BSSCs) that can be used to identify sleep stages in preterm infants born <37 weeks gestational age. However, they operate different criteria to define sleep stages, which limits the comparability and reproducibility of research on preterm sleep. This scoping review aims to: 1) identify and elaborate on existing neonatal BSSCs used for preterm infants, 2) examine the reliability and validity of these BSSCs, and 3) identify which criteria are most used for different ages, ranging from 23 to 37 weeks postmenstrual age at observation.
Methods: To map the existing BSSCs, PubMed, EMBASE and Cochrane were searched for studies using a BSSC to identify sleep stages in preterm infants.
Results: In total, 36 BSSCs were identified with on average five item categories assessed per BSSC, most frequently: eyes, body movements, facial movements, sounds, and respiratory pattern. Furthermore, validity and reliability of the BSSCs were tested in less than half of the included studies. Finally, BSSCs were used in infants of all ages, regardless the age for which the BSSC was originally developed.
Conclusions: Items used for scoring in the different BSSCs were relatively consistent. The age ranges, reliability, and validity of the BSSCs were not consistently reported in most studies. Either validation studies of existing BSSCs or new BSSCs are necessary to improve the comparability and reproducibility of previous and future preterm behavioral sleep studies.

1. Introduction

Each year, an estimated 10% of all live births are preterm (born before 37 weeks of gestation) [1]. These infants are born during a very critical period of brain development [2]. Not surprisingly, preterm infants are at risk for adverse long-term neurodevelopmental outcomes [2]. Sleep is consistently found to have a protective effect on brain development in preterm infants [3–7]. Hence, promoting sleep in infants born preterm should be one of the primary concerns in the neonatal intensive care unit (NICU).

In preterm infants, sleep behavior is often classified into three stages, which are categorized according to the level of behavioral activity. Periods of seemingly restless behavior are often classified as ‘active sleep’ (AS), which is the preterm equivalent of rapid eye movement sleep (REM-sleep). AS makes up 40–60% of the total time preterm infants spend sleeping [8]. This stage is considered important for the endogenous stimulation of sensorimotor processing areas of the brain, facilitating activity-dependent development [6,9,10]. More tranquil behavior is often called ‘quiet sleep’

Abbreviations: AS, Active sleep; QS, Quiet sleep; IS, Intermediate sleep/Indeterminate sleep; BSSC, Behavioral sleep stage classification method; EEG, Electroencephalography; PSG, Polysomnography.

* Corresponding author. Department of Neonatology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Lundlaan 6, 3584, EA, Utrecht, the Netherlands.
E-mail address: j.dudink@umcutrecht.nl (J. Dudink).

[1] Authors contributed equally.

https://doi.org/10.1016/j.sleep.2022.01.006

(QS), also known as the preterm equivalent of non-rapid eye movement sleep (NREM-sleep). QS is thought to induce experience-dependent synaptic remodeling via repetitive, synchronized activity within neuronal pathways [9,10].

Often an intermediate or indeterminate sleep stage (IS) is included in the classification of sleep stages. IS can be considered either a transitional sleep stage (intermediate sleep) [11,12] or can be classified when an infant's behavioral sleep stage is unclear (indeterminate sleep) [13,14]. Next to the three sleep stages, the behavioral state 'wake' exists, which can be further subdivided into multiple stages or behaviors, such as alert, drowsy, fussy, or crying [15]. The full impact of being awake versus being asleep on preterm brain maturation is still largely unclear in humans. However, animal studies show that suppression of sleep leads to impaired brain development [4,16–18]. Because of the consistent findings of the protective ability of sleep on brain development in humans [3–7] and the findings in animal studies [4,16–18], sleep is frequently considered the most important driver of early brain development and maintenance [10,20].

It is important to note that sleep staging is often done on a small time window. However, sleep-wake states also show a 24-h our pattern. Both regularity and cyclicity of this pattern provide an important context exhibiting quality and quantity of sleep behavior [19]. Considering the clear connection between preterm sleep and development, sleep stages and the 24-h sleep-wake rhythm can be used as a target for interventions [21–23] and as a biomarker indicating early developmental advancement, or dysmaturation [24]. Therefore, it is important that sleep stages are classified in a reliable and valid manner and that their monitoring is barely obtrusive. Furthermore, it is important that assessment is easy, accessible, and applicable in all imaginable clinical circumstances. Of all methods to assess sleep in preterm infants (for a review see Ref. [25]), behavioral observations seem to best meet these criteria.

Over the last decades, a variety of observational, behavioral sleep stage classification methods (BSSCs) in preterm infants have been developed and refined, roughly using the same sleep stages as described above. However, there are differences in definitions and in methods of validation used in various BSSCs. Also, BSSCs use a large variety of items to classify sleep stages. The lack of a gold standard for BSSCs may result in differences in methods and findings between preterm sleep studies, which could complicate replication of, or comparison between, these studies.

Therefore we performed a scoping review to determine the scope of the body of literature on this topic and give clear indication of the volume of literature and studies available as well as an overview of its focus [26]. Scoping reviews are useful for examining emerging evidence when it is still unclear what other, more specific questions can be posed and valuably addressed by a more precise systematic review [27]. Scoping reviews can report on the types of evidence that address and inform practice in the field and the way research has been conducted. This scoping review aims to identify and elaborate on existing neonatal BSSCs used in preterm infants (<37 weeks postmenstrual age (PMA) at times of assessment). Their reliability and validity will also be examined. Finally, this review aims to specifically identify which criteria are most used for each age group, ranging from 23 to 37 weeks PMA at observation.

By providing an overview of existing BSSCs and elaborating on each of them, the comparability and replicability of sleep studies will improve. Furthermore, researchers can consult this overview to choose an appropriate BSSC. All together, by providing an overview of the BSSCs, identifying the most useful scoring items, and reference to the reliability and validity of each BSSC, it will be 1) easier for researchers to develop new BSSCs for different samples – eg, based on age, 2) clear which BSSCs are not only seated in habitual practice but also rely on valid and reliable research, and 3) easier to replicate studies using BSSCs in preterm infants.

2. Methodology

2.1. Design

This scoping review existed of a three-step plan. Firstly, a systematic search of peer-reviewed scientific literature was executed to find all experimental studies using BSSCs to assess sleep in preterm infants. Secondly, from these experimental studies, the BSSC that was used was traced back (by authors AB and CS). These BSSCs were included for further analysis. Furthermore, if the experimental study used a new in-house developed BSSC, this BSSC was included for further analysis. Thirdly, the reference lists of all included BSSCs were checked to identify any additional studies introducing a BSSC (by AB, CS and EG). Thus, in the end three categories of studies were searched for in this scoping review:

1) Studies that introduced a newly developed BSSC. These studies were found using the systematic search string;
2) Studies using an existing BSSC to assess sleep in preterm infants. These studies were also found using the systematic search string;
3) Other literature on the BSSCs that were used by studies of category 1. These BSSCs were found using the reference lists of the studies of category 1.

The BSSCs (category 1 and 3) were evaluated on three features: 1) the division of sleep stages, 2) the exact criteria (ie, items) used to classify each sleep stage, and 3) the validity and reliability when used in preterm infants. The principles of the 'Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews' (PRISMA-ScR) were adhered to in order to facilitate complete, transparent, and consistent reporting of the literature [28].

2.2. Information sources and search strategy

First, an explorative search on Cochrane was performed to map any existing scoping reviews on the use of BSSCs for preterm infants. The search with terms based on keywords including 'behavior', 'sleep', and 'preterm', yielded nine hits but none of these were on the use of BSSCs in preterm infants. After concluding that no scoping review on this topic existed, a systematic literature search was performed on 21-12-2020 in PubMed and on 28-01-2021 in EMBASE to ensure that all BSSCs used for preterm infants are included in this review. An update search in both databases was performed on 27-08-2021. The search string (see Appendix A) based on the same keywords ('behavior', 'sleep' and 'preterm') was created in collaboration with a content expert in preterm infant research (JD) and a systematic review expert (AH). To their knowledge, the experts selected all synonyms of the keywords used in scientific literature to optimize the thorough search and capture relevant articles. In addition, a librarian proficient in systematic search techniques provided her professional opinion regarding the search strings.

This scoping review was performed according to the JBI scoping review guidelines (https://jbl.global) and focused on published studies and scientific-based BSSCs, as these BSSCs had been available for usage by the scientific community. Accordingly, no active search was conducted outside of PubMed, EMBASE and Cochrane, since literature obtained via such a search would not have been available to all researchers interested in sleep and preterm infants.

2.3. Inclusion criteria and exclusion criteria

To be eligible for inclusion in this review, studies had to include, but were not necessarily limited to, preterm infants who were <37 weeks PMA. Furthermore, infants had to be admitted to a neonatal ward at the time of observation. Moreover, only studies that used BSSCs were included. More specifically, this comprised:

- Studies that introduced a new BSSC for preterm infants.
- Studies that used an existing BSSC, but slightly modified it. A clear description of the modification was required.
- Studies that used an existing BSSC, in its original form.

Furthermore, studies were only included if two or more sleep stages were distinguished by the BSSC during the scoring session of that study. If sleep stages were merged later in the process, eg for the purpose of data analysis, the study could still be included, as long as it was clear that the sleep stages were initially distinguished during data collection. Studies that used both a BSSC and one or multiple technological approaches, including electroencephalogram (EEG), actigraphy, video analysis or polysomnography (PSG), were only included when identification of sleep stages could be performed through independent use of the observational characteristics (ie, the BSSC), either during real time observations or by watching videos of the preterm infant. Finally, a study was only included if the criteria used to assess the sleep stages were clearly specified.

Studies were excluded if they included non-human participants. Also, case studies, systematic reviews, meta-analyses, and publications (including conference abstracts) without the full text available were excluded. Furthermore, all studies in languages other than English or Dutch were excluded.

2.4. Study selection

Following the search, all identified studies were uploaded into Rayyan QCRI (Qatar Computing Research Institute, Data Analytics [29]) and duplicates were removed. Subsequently, studies were screened on inclusion/exclusion criteria by two independent reviewers (AB and CS), first on titles and abstracts, later on full texts. From all included studies, the BSSCs were extracted. All BSSCs were double-checked by at least two independent researchers (AB, CS, and/or EG). During all stages of the study selection, any disagreements were resolved by discussing them with a third reviewer (either SV, AH or EG).

2.5. Data extraction and synthesis

After mapping which BSSCs were used in the included studies, the online databases of Utrecht University were searched to collect these BSSCs. For BSSCs that could not be found online, the private literature collection of the Wilhelmina Children's Hospital of the University Medical Center Utrecht was consulted. If the BSSCs could not be found either online or in the private collections, the corresponding author was contacted with a request to share their work.

The following information was extracted from the included studies and its extracted BSSCs: 1) the total number of included studies using a particular BSSC, 2) the total number of patients scored using that BSSC, 3) the validity and reliability of the BSSC (assessed by either the original authors, by studies using the BSSC, or both) and 4) the age ranges for which the BSSC was originally developed. All information was extracted into chart tables. For a detailed overview of the information in these tables, see the protocol of this scoping review [30].

3. Results

3.1. Descriptives

The database searches yielded 3961 citations of studies. To request access to known but unavailable studies, authors were contacted. Of the 138 requests sent, 14 authors replied, and 13 sent the requested study. Six of the 13 studies were included (Appendix B1, references with an asterisk). Following assessment of eligibility, 147 studies were included for qualitative data analysis (Appendix B1, sections one and two for an overview of all included studies). Subsequently, 47 BSSCs were identified (Appendix B2, sections one and two for an overview of all BSSCs). Of the 147 included studies, 11 were identified as BSSC (Appendix B2, section one, references with an asterisk); the other 25 BSSCs were extracted from the experimental studies. For an overview of the selection process, see the PRISMA Figure (Fig. 1).

After the initial selection, a continuous assessment and reassessment of the studies and the BSSCs took place, in order to extract all details, make corrections and add any missed information. Accordingly, several studies and BSSCs were excluded. Of the 47 identified BSSCs, eight were not available in full text [31–38]. Two BSSCs could not be clearly identified [22,39] – eg, the BSSC was only mentioned in a general manner: “Prechtl was used for sleep classification”, [22]. Therefore, it was unclear which particular BSSC was used. Finally, one BSSC was referred to as a BSSC, but physiological parameters seemed essential for classification, which was an exclusion criterium for this scoping review [40]. Therefore, these eleven BSSCs were excluded from further analysis (Appendix B2, section two), resulting in 36 remaining BSSCs (Appendix B2, section one). Ten of the 147 included studies referred to excluded BSSCs (Appendix B1, section two) and were therefore excluded from further analysis.

The remaining 137 included studies (Appendix B1, section one) were published between 1980 and 2021. Study designs included randomized control trials, pre-post studies, cohort studies, and case–control studies. Sample sizes in the studies ranged between five and 147 infants, with a combined total of $n = 8847$ infants. The age of the preterm infants ranged from 23 to 37 weeks GA at birth and from 23 to 37 weeks PMA at the time of observation. Some studies included term infants, but did not specify the ratio of preterm and term infants. Therefore, the combined total of $n = 8847$ partly includes term infants [15,44,47].

The remaining 36 BSSCs were published between 1959 and 2013. For a detailed overview of the characteristics of the BSSCs and the number of included studies using the BSSCs, see Table 1. Of the 36 BSSCs, Brazelton's sleep stage assessment BSSC and Prechtl's BSSC were used most frequently, respectively in 38 and 23 included studies (Table 1). Note that both the Neonatal Behavioral Assessment Score (NBAS; referred to in 21 studies) and Newborn Individualized Developmental Care and Assessment Program (NIDCAP; referred to in 17 studies) fall under the Brazelton's BSSC, as Brazelton's sleep stage assessment is part of both the NBAS and the NIDCAP. For a detailed overview of all BSSCs using Brazelton's BSSC, see Appendix B3.

3.2. Item overview

Throughout the 36 BSSCs, five categories of scoring items frequently recurred: eyes, body movements, facial movements, sounds, and respiration pattern. The median of item-categories used by the BSSCs to classify behavioral states is five. Thoman et al. [41] used the most with seven categories of items. Stefanski et al. [42] used the least with two categories of items.

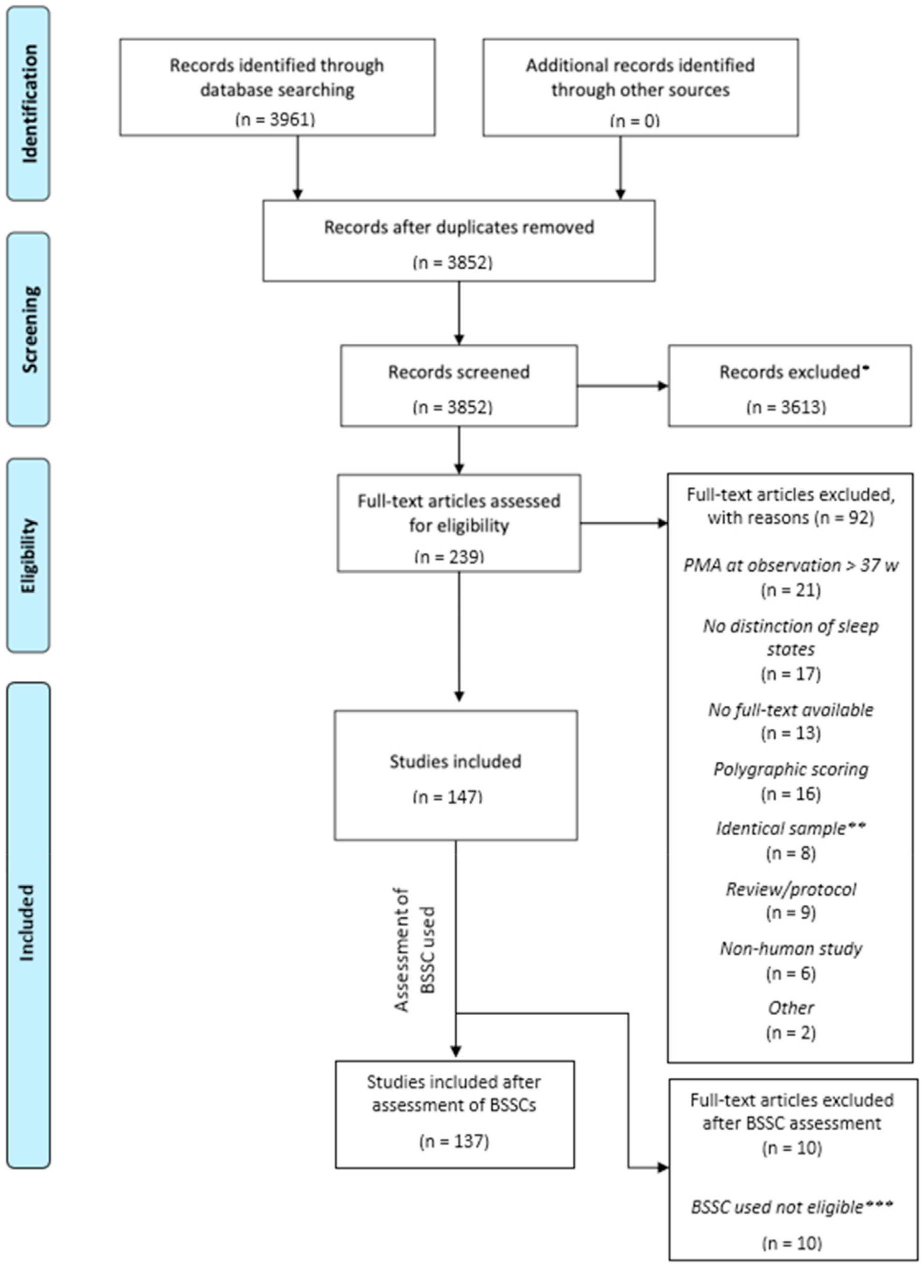

Fig. 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR). *Of the 3613 studies excluded at screening, the full texts of 96 were unavailable, and the abstracts did not contain enough information to meet the inclusion criteria. **Eight studies were excluded because they published several studies about the exact same dataset. The inclusion of multiple studies using the same dataset would have distorted the data regarding how often a BSSC had been used with the preterm population. ***After thorough analysis of the BSSCs extracted, eleven turned out to be ineligible for inclusions. Retrospectively, studies referring to and making use of these ineligible BSSCs only (n = 10) were excluded from further data analysis.

The five most frequently used categories were eyes (eg, eye movements), body movements, facial movements, sounds, and respiration pattern (for a detailed overview, see Table C.1, Appendix C). Within these categories, the behavioral stages QS, AS, wake, and the non-specific stage 'sleep' showed different trends. According to the BSSCs, open eyes were seen most dominantly in wake, whereas closed eyes were dominant in the sleep stages. In addition to closed eyes, AS was linked to rapid eye movements (REM).

The characteristics of body movements seemed to show a more uniform trend over the behavioral states. However, low muscle tone was connected more often to AS than QS or wake. In QS, either no or reflexive body movements were most characteristic, and a high activity level was most often found during wake. Additionally, regarding facial movements, preterm infants showed primarily rhythmic movements during QS, whereas during AS, many non-rhythmic facial movements were apparent.

Furthermore, sounds were clearly different for each behavioral state. Only releasing sounds, including sobs and sighs, could be classified during QS. During AS and wake, the infants also made other noises, with distressed noises, including crying, being most prominent in wake. Finally, the respiration pattern was characteristic of the sleep stages. The most common distinction was regular respiration in QS and irregular respiration in AS.

Table 1
An overview of the characteristics of the 36 BSSCs extracted from the search or included studies. The name of each BSSC is based on either a) the name assigned to the BSSC by the original author; or b) the name of the first author of the original publication of the BSSC. The order of the BSSCs is based on the number of studies referring to the BSSC, with numbers in descending order.

BSSC	N studies referring	N patients	Validity		Reliability		Age range (in authors' words)
			Original	Other	Original	Other	
Brazelton's BSSC (combined)	38	1653	no	yes	yes	yes	Preterm & full-term
	NBAS (1973, 1977, 1984, 1995; NNNS, 2004)[a]	21	836				
	NIDCAP (1981, 1984, 1995; APIB, 1982; Als in Goldson, 1999)	17	817				
Prechtl (1969,1974)~	23	587	no	yes	no	yes	Full-term
Stefanski (1984)[a]	21	1023	no	no	yes	yes	PCA 30–42 weeks
ABSS[b]	13	508	no	no	no	yes	M = 30 weeks GA at birth + 24 days old; Thus, 33.4 weeks. PMA
Thoman (1975, 1976)~	13	484	no	no	no	yes	Full-term
Holditch-Davis (1990)[b]	9	627	no	yes	yes	yes	29–39 weeks PMA
Thoman (1990)[b]	8	542	yes	no	yes	yes	–
Anders (1971)[a]	8	241	no	no	no	yes	Full-term
Wolff (1966)[b]	6	255	no	no	yes	yes	Day 4 PNA; GA/PMA not specified
Korner (1972)[b]	4	176	no	no	yes	yes	Full-term
Holditch-Davis (2004)[b]	3	217	no	yes	yes	yes	M = 28.8 weeks GA at birth, weekly until discharge/44 weeks PMA
Emde (1969a, 1969b)~	3	70	yes	no	yes	yes	GA 38–42 weeks
Mercuri (1995)[b]	2	32	no	no	no	no	<36 weeks PMA
Watanabe (1992)~	2	10	no	no	no	no	24–46 weeks PCA
Liaw (2012, 2013)[b]	2	140	no	no	no	no	(GA) 26–37 weeks born + 2–28 days postbirth observation
Doussard-Rossevelt (1996)[b]	1	62	no	no	no	no	PCA 33–35 weeks
Loewy (2013)[b]	1	272	no	no	no	no	M = 29.57 weeks GA, every day until M = 22 days PNA
Parmelee (1972)[b]	1	14	no	no	no	yes	32 weeks PMA - 8 months past term
Thoman (2001)	1	97	no	no	no	no	–
Wolff (1959)[b]	1	4	no	no	yes	no	0–5 days PNA; GA/PMA not specified
Giaganti (2002)[b]	1	8	no	no	yes	no	30–35 weeks PCA observation, (GA) 28–35 weeks born
Brandon (2005)[b]	1	56	no	no	yes	no	Observation during week 32 or 36 PMA, GA 24–29 weeks born
Ingersoll & Thoman (1999)[b]	1	95	no	no	yes	no	33–35 weeks CA, GA 26–31 weeks

Behavioral sleep stage classification method (BSSC); Anderson Behavioral State Scale (ABSS); Neonatal Behavioral Assessment Scale (NBAS); Neonatal Intensive Care Unit Network Neurobehavioral Scale (NNNS); Newborn Individualized Development Care and Asssessment Program (NIDCAP); Assessment of Preterm Infants' Behavior (APIB); number of (N); postconceptional age (PCA); mean (M); gestational age (GA); postmenstrual age (PMA); postnatal age (PNA); conceptional age (CA).

[a] N patients include some full-term infants (if the study included both preterm and full-term infants, and it was not specified in the study how many infants were preterm and how many full-term).

[b] N patients include the sample used during development of BSSC. In some BSSCs the N patients was not specified. ~ N patients include full-term infants and the original sample. Note: The NBAS and the NIDCAP both fall under Brazelton's BSSC and encompass one and the same BSSC. We identified multiple identical editions of both the NBAS and the NIDCAP. For an overview of all BSSCs falling under the category 'Brazelton's BSSC', see Appendix B3. Moreover, the BSSCs of Emde, Liaw, Prechtl, and Thoman each have two identical editions of the same score and are also combined.

Additionally, some BSSCs included several other categories for classifying sleep stages. These items involve breathing technique (eg, abdominal breathing) [41,42], response to stimulation [61], and the color of the face [12,53]. However, these items were seldom used, thus are not considered in the final analysis.

3.3. Individual scores: age ranges and items

Between the individual BSSCs that were identified in this review, there were some notable differences in the age ranges and items. The BSSCs were developed for a broad range of infants, ranging from 24 weeks PMA until full-term age. In total, 14 of the 36 BSSCs were originally developed for preterm infants. However, most of these BSSCs were originally developed for infants >30 weeks PMA. Watanabe and Liaw's BSSCs [14,41,42] were the only BSSCs developed for a preterm population including very (<30 weeks GA) to extremely (<28 weeks GA) preterm infants. However, both BSSCs were used only by their own research group. Additionally, Watanabe et al. [14] used their BSSC to study a sample of preterm infants >30 weeks GA, which means their own experimental sample did not include very to extremely preterm at the time of observation.

As derived from the included experimental studies, the BSSCs most frequently used in an extremely preterm sample were Brazelton's (NBAS/NIDCAP; eight times) [44, Appendix B3], Prechtl's (seven times) [37], Anderson's (ABSS; six times) [53], and Thoman's (five times) BSSC [11]. Also, the BSSCs most frequently used in a very preterm sample were Prechtl's (nine times) [37] and

Brazelton's (NBAS/APIB/NNNS; five times) BSSCs [44, Appendix B3], followed by Anders' (four times) [15] and Holditch-Davis' BSSCs (three times) [48].

When comparing the BSSCs used for very to extremely preterm infants, some differences were evident between the BSSCs originally developed for this population (ie, Watanabe and Liaw [14,41,42]) and the most frequently used BSSCs in this population (ie, Brazelton and Prechtl [43,44, Appendix B3]. First, both Watanabe and Liaw's BSSCs did not take facial movements into account when assessing sleep stages. Also, Liaw's BSSC was more liberal than others in which items can occur during QS. For example, according to Liaw's BSSC, fussing can be assigned to QS, while the other BSSCs used in preterm infants <30 weeks PMA only acknowledged fussing occurring during wake. Furthermore, Liaw's BSSC identified fewer movements that can occur during AS in comparison to the other BSSCs. For example, Liaw's BSSC states that AS is characterized by low motor activity with low muscle tone, while both Brazelton's and Prechtl's BSSCs included multiple categories of movements that may occur during AS, including startles and twitches. Lastly, of all BSSCs identified in this review, most included respiration rate (eg, fast in AS), and some also included heart rate (eg, regular and slow in QS, and irregular in AS) for the categorization of sleep stages. However, neither Watanabe's nor Liaw's BSSC used the respiration rate and heart rate for the categorization of sleep stages.

On the other hand, there were similarities between the BSSCs. All BSSCs took both REM as characteristic of AS and no eye movements as characteristic of QS. Also, rather than the respiration rate, the respiration regularity was taken into account in each BSSC. Specifically, irregular respiration was listed as characteristic of AS, while regular respiration was listed as characteristic of QS.

3.4. Validity and reliability

Of the 36 included BSSCs, reliability had been assessed for 15 BSSCs [11,12,15,43–55, Appendix B3], and six BSSCs had been validated. For a detailed overview of the reliability and validity assessments, see Tables D.1 and D.2 in Appendix D respectively. The validated BSSCs included two that had been validated by the original author [11,49,50] and four that had been validated by other research groups [43,44,48,55, Appendix B3]. Within these studies, mainly content validity [43,44,48, Appendix B3] and criterium validity [44,49,50,55, Appendix B3] were assessed. Construct and predictive validity were checked in one BSSC [11]. Also, face validity was assessed in one BSSC [43]. In all studies, few details were given regarding (parts of) the validation process. Only two studies provided extensive descriptions of the way in which predictive [11] and criterium [55] validity had been assessed. Finally, about one BSSC [47] it was mentioned that 19% of REMs detected by an electrooculogram (EOG) were not detected during observation, providing an indication of instrument validity of this specific item.

The reliability of three BSSCs had been checked by the original authors [49–52]; the reliability of five BSSCs had been checked by other authors [15,43,45,46,53], and for seven other BSSCs, reliability had been checked by both the original and other authors [11,12,44,47,48,54,55, Appendix B3]. The reliability of the BSSC as a whole was most commonly calculated using a Cohen's or Fleiss' kappa. This was done for eight BSSCs [11,15,43,44,46,48,53,55, Appendix B3]. These eight BSSCs had a kappa ranging between 0.31 and 0.96. For five studies [43,47,48,51,52], a kappa per sleep stage was calculated, ranging between 0.61 and 0.93 for AS, between 0.41 and 0.86 for QS, and between 0.823 and 0.90 for wake. For one BSSC, reliability measures for separate items were calculated, resulting in a kappa between 0.82 and 1.00 for body movements [46].

The other included studies assessing reliability did not explicitly state that a kappa had been calculated. The outcomes of the reliability assessments were frequently referred to as “interrater reliability”, “percentage interrater agreement” (commonly a correlation coefficient), or “interrater”. Generally, these non-specific measures of reliability were higher than the commonly used “kappa”. More specifically, they ranged between 77.6% and 100% or 0.90–0.983.

4. Discussion

The aim of this scoping review was to identify which existing neonatal BSSCs are commonly used for preterm infants (<37 weeks PMA) and to analyze their reliability and validity for this age group. Additionally, this review aimed to specifically identify which scoring items were most used for each age group, ranging from 23 to 37 weeks PMA at observation. The search resulted in 47 BSSCs, of which 36 were eligible for analysis. The findings showed that the items used for sleep assessment using these BSSCs were rather comparable. Moreover, reliability and validity of these BSSCs were not consistently tested in most studies.

4.1. Age ranges and items

Most of the BSSCs studied were very comparable, with five frequently recurring categories of scoring items: eyes, body movements, facial movements, sounds, and respiratory pattern. The review found that scoring items in the categories of eyes and respiration pattern were mostly consistent across all BSSCs for each sleep stage. However, literature shows that very to extremely preterm infants may have a faster respiration rate in QS than in AS, which seems to be reversed in older preterm infants [56]. Furthermore, in contrast to older preterm and full-term infants, preterm infants <32 weeks PMA may not have developed cardiorespiratory coupling yet. From the current review it became clear that also BSSCs developed for full-term infants are used in very to extremely preterm infants. This means that invalid conclusions may have been drawn from observations of cardiorespiratory behaviors. Therefore, it is of utmost importance that only BSSCs that are validated for very to extremely preterm infants are used for this specific age group.

Facial movements were not included in the BSSCs that were developed for extremely preterm infants [14,41,42]. However, it is not explained in these BSSCs [14,41,42] why the decision was made to exclude facial movements. A possible explanation could be the difficulty in recognizing facial movements related to sleep in the very to extremely preterm samples. These infants often wear a face mask for respiratory support, which restricts the visibility of mouth movements. Facial movements were however considered in most other BSSCs that were also used for extremely preterm infants [43,44, Appendix B3]. This gives the impression that facial movements – if visible – can be used to classify sleep stages for very to extremely preterm infants but are not validated yet for this sample. Therefore, further research should be conducted on the validity of using facial movements for classification of sleep stages in a very- to extremely preterm patient group.

In addition to the five most frequently used categories, some BSSCs included several other categories for classifying sleep stages. These items include breathing technique, response to stimulation, and the color of the face. These categories were not included in the five most frequently used categories, given their scarce appearance in scores. Nevertheless, in a study performed on cats, Joan et al. [62] suggested that the motor control of the diaphragm is different during REM sleep than in other states. Since breathing movements seem to be connected to different sleep stages in other mammals, it

Table 2
All factors that need to be considered and specified when using or validating a BSSC.

Factors for consideration	Specifics to discuss in literature
Patients	Specify the age of the patients. Specify the sleep condition of the patients regarding their visibility (covered eyes/blanket on the body/cover on the incubator/ sleeping position).
Observers	Specify the experience and training of the observers.
Observation	Specify the length of the full observation as well as the time of day/night at which it is conducted. If the observations were long/nightly, mention how observer fatigue was dealt with. Mention whether and when the observations were interrupted for care by hospital staff. Note the length of each epoch and how the observers were keeping track of the epochs. Specify how the observers were taking notes of the observations (eg by using a standardized form).
Items	Specify the different items that were assessed during each epoch. Describe the items in detail to prevent misinterpretations.
Validation	Specify if and how the BSSC has been validated. Include the name of the statistical test and discuss the results. Content and construct validity are both important indicators that can be assessed , eg by reducing the number of items and seeing if the same construct is measured, through assessing reliability between the original and the reduced scale.
Reliability	Specify if and how the BSSC has been tested for its reliability. Include the name of the test and discuss the results briefly. It is recommended to use a well-known measure for reliability, so the scale can be compared with other scales. The most used way to measure reliability is a Cohen's or Fleiss' kappa. Reliability can be assessed between observers, but also within observers (eg by letting an observer assess the same video twice). It is important to assess reliability for a longer period of time and/or assess reliability multiple times to represent a full observation.

Behavioral sleep stage classification method (BSSC).

could be of interest to focus future sleep research on breathing techniques in preterm infants. Also, response to stimulation possibly indicates sleep stages, as demonstrated by Kuhn et al. [63]. They found that preterm infants respond to loud, high-frequency, and artificial NICU sounds, which can affect their sleep quality [63]. Nevertheless, before these three items can be considered for implementation in future BSSCs, more research should be conducted.

The age for which a BSSC is developed does not seem to be the most important criterium for choosing which BSSC to use for a preterm sample. More specifically, several included experimental studies used a BSSC that was originally developed for full-term infants to assess behavioral states in a preterm sample [15,43]. This contradicts literature suggesting a significant difference in behavior between the very to extremely preterm period and older preterm infants [10,64]. It is important to mind that high validity may not be assured if a study uses a BSSC in a different age sample than it was developed for.

On another note, some included studies used a PSG-based method to assess behavioral states, but only used the behavioral items of this score. In some cases, these behavioral items were used as a stand-alone scoring method [57–60] even though the authors that had developed the score did not communicate this possibility in their original publication [40]. In this review, studies that used only the behavioral criteria of PSG scores were excluded from further data analysis, in case there has not been a validation of these criteria as stand-alone score or any other indication of the possibility to use these behavioral criteria as such. However, Anders' BSSC was an exception [15]. Anders and colleagues clearly elaborated on the behavioral criteria of their PSG score, while acknowledging that these behavioral criteria were sufficient to be used as a stand-alone BSSC. However, before using these behavioral criteria of Anders as stand-alone BSSC in the future, it is important that these behavioral criteria are validated as such.

4.2. Validity and reliability

The reliability assessments of the original BSSCs raise some concerns. Most assessments refer to an interrater agreement (Table D.1, Appendix D); however, this is often with little detail about the assessment performed. Moreover, the timing of the reliability assessment could make it invalid, as subjectivity or observer fatigue may introduce biases in the assessment. For example, when the behavioral state classification of one researcher at 15 minutes before the end of a three-hour observation is tested against the classification of the first 15 minutes of the observation period of a second researcher [11]. In this case the first observer is fatigued due to the long observation, whereas the second observer is not fatigued. Furthermore, the first observer has had almost three hours to get to know the infant's behavior, whereas the second observer does not have this knowledge. Moreover, a reliability test conducted solely at the beginning of a lengthy observation period may not represent the reliability of the full period. Reliability should therefore be tested during a longer observation period and/ or several times over a longer observation period, and this should be clearly described in detail in the published results. Finally, reliability is preferably also retested − eg, by observing a video and assessing it twice. This will show how consistent observers are over time.

Another timing related factor that may influence the validity of the score, is the time of day or night at which the scoring is performed. Often, no recommendations are made by the BSSCs regarding the timing of observations. Therefore, it remains unclear if time of day may influence validity of the observations. This unknown factor potentially limits the generalizability between observations at different time points. To simplify the process of finding the right BSSC for a specific research, it would be useful for researchers to know during which time period a BSSC was validated.

Additionally, this review found that validity assessments were rarely performed on or described in the BSSCs (Table D.2, Appendix D). If validity assessment was performed, it was barely explained in detail how validity was assessed [11,32,44,48], [Appendix B3]. In one case, the BSSC was validated for full-term infants but had also been used with preterm samples [43]. Interestingly, Anders' BSSC was used multiple times for (very to extremely) preterm samples. However, it was specifically pointed out in the original study that their scoring method, developed for full-term infants, was neither appropriate, nor validated, for preterm infants [15]. Of the BSSCs that were developed for preterm infants, only half of them were assessed for either reliability or validity. Thus, both validity and reliability of a BSSC do not necessarily seem to be considered when selecting a BSSC. This raises the question on which criteria the selection of a BSSC is based instead.

4.3. Limitations and further research

One limitation of this study may be the summarizing of items in the BSSCs to get a clear overview. This may have resulted in a biased overview of categories due to different names of characteristics, that also might have been slightly different in nature. For example, the items that indicated a reflex (eg, jitter, twitch, startle) were merged to the item 'reflexive movement'. Additionally, it was barely described by the BSSCs how items were precisely assessed. Frequently, a list of items was provided by the BSSC, without providing a detailed explanation about what each item meant. This led to the necessity to make assumptions, in order to summarize which items and categories of items were used by all BSSCs. Although the underlying thought of all BSSCs is the same, most BSSCs have at least one specific feature that is different from the other BSSCs. Such features may be an extra category of items (eg, breathing technique) or more specificity regarding the items (eg, more sub-items stating which specific reflexive movements can occur). In addition, of some BSSCs multiple versions were available. The differences between new and older versions of a BSSC were often not elaborated on by its authors in the publication of the new BSSC. In some cases, the different versions of the BSSC did not seem to differ, thus were merged for data analysis in the present scoping review.

The development of new technology to monitor preterm infants' sleep stages is an ongoing process. The benefits and pitfalls of existing methods, among which EEG and PSG, are extensively described in multiple reviews [25,65]. However, it is likely that even more unobtrusive methods will be developed in the future. Validation of such noticeable technological improvements towards sleep research is highly important and could lead to new advances in this research area. Having validated BSSCs to compare, validate and compliment new scoring methods could stimulate the progress in the research field, which in turn could lead to improved sleep protocols in the clinics. The creation of an unobtrusive gold standard containing descriptions of how to measure sleep behavior in preterm infants is an important next step in the process. This validated gold standard should include information on multiple subjects as listed in Table 2.

5. Conclusion

To conclude, 14 out of 36 BSSCs used for preterm behavioral sleep stage scoring were originally developed to assess sleep stages in preterm infants. The BSSCs varied in the items they used for scoring different sleep stages. However, some items, particularly in the eyes and respiratory pattern categories, showed several similarities across the BSSCs. Validity testing was not performed or described in a standardized manner, which renders the conclusions based on some BSSCs questionable. To avoid unclarity in future studies and to strengthen present preterm behavioral sleep studies, either existing BSSCs should be validated (and adjusted) properly, or a new gold standard BSSC should be developed for sleep research in preterm infants. This would promote sleep research on vulnerable preterm infants. Eventually, more knowledge on preterm infants' sleep organization will lead to improved health care treatment as sleep-wake cycles can be used for planning elective care. Furthermore, sleep quality and quantity assessment could serve as an early biomaker for typical brain development. Ultimately, sleep optimization during their NICU stay will give infants a better chance on healthy brain maturation.

Funding source

The authors acknowledged the financial support of EU H2020 MSCA-ITN-2018: INtegrating Functional Assessment measures for Neonatal Safeguard (INFANS), funded by the European Commission under Grant Agreement #813483.

Acknowledgements

Also, we would like to thank Charlotte Teunis and Roos Bos for their preliminary literature search and locating original articles and textbook copies during their research internship.

Conflict of interest

We declare no competing interests.

The ICMJE Uniform Disclosffigure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: https://doi.org/10.1016/j.sleep.2022.01.006.

Appendix A-D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sleep.2022.01.006.

References

[1] Beck S, Wojdyla D, Say L, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. Bull World Health Organ 2010;88(1):31–8. https://doi.org/10.2471/BLT.08.062554.
[2] Bennet L, Walker DW, Horne RSC. Waking up too early – the consequences of preterm birth on sleep development. J Physiol 2018;596(23):5687–708. https://doi.org/10.1113/JP274950.
[3] Ednick M, Cohen AP, McPhail GL, et al. A review of the effects of sleep during the first year of life on cognitive, psychomotor, and temperament development. Sleep 2009;32(11):1449–58. https://doi.org/10.1093/sleep/32.11.1449.
[4] Mirmiran M, Scholtens J, Van de Poll NE, et al. Effects of experimental suppression of active (REM) sleep during early development upon adult brain and behavior in the rat. Dev Brain Res 1983;7(2–3):277–86. https://doi.org/10.1016/0165-3806(83)90184-0.
[5] Mirmiran M, Van de Poll NE, Corner MA, et al. Suppression of active sleep by chronic treatment with chlorimipramine during early postnatal development: effects upon adult sleep and behavior in the rat. Brain Res 1981;204(1):129–46.
[6] Roffwarg HP, Muzio JN, Dement WC. Ontogenetic development of the human sleep-dream cycle. Published by American Association for the Advancement of Science Stable URL :; 1966. p. 604–19. vol. 152, no. 3722, https://www.jstor.org/stable/1718980.
[7] Hubel DH, Wiesel TN. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. J Physiol 1970;206(2):419–36.
[8] Curzi-Dascalova L, Peirano P, Morel-Kahn F. Development of sleep states in normal premature and full-term newborns. Dev Psychobiol 1988;21(5):431–44.
[9] Peirano P, Algarín C, Uauy R. Sleep-wake states and their regulatory mechanisms throughout early human development. J Pediatr 2003;143(4 SUPPL):70–9. https://doi.org/10.1067/s0022-3476(03)00404-9.
[10] Knoop MS, de Groot ER, Dudink J. Current ideas about the roles of rapid eye movement and non–rapid eye movement sleep in brain development. Acta Paediatr Jan. 2021;110(1):36–44. https://doi.org/10.1111/APA.15485.
[11] Thoman EB. Sleeping and waking states in infants: a functional perspective. Neurosci Biobehav Rev 1990;14(1):93–107. https://doi.org/10.1016/S0149-7634(05)80165-4.
[12] Wolff PH. The causes, controls, and organization of behavior in the neonate. Psychol Issues 1966;5(1):1–105.
[13] Thoman EB. Early development of sleeping behaviors in infants. In: Norman ER, editor. " in Abberant development in infancy. Abingdon-on-Thames. Oxfordshire, England, UK: Routledge; 2018. p. 123–38.
[14] Watanabe K. Electroencephalogram and sleep-cycle patterns. In: Eyre JA, editor. " in the Neurophysiological examination of the newborn infant. New York: Cambridge University Press; 1992. p. 11–47.
[15] Anders T, Emde R, Parmelee Jr AH. A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants. Los Angeles: UCLA Brain Information Service; 1971.
[16] Shaffery JP, Oksenberg A, Marks GA, et al. REM sleep deprivation in monocularly occluded kittens reduces the size of cells in LGN monocular segment. Sleep 1998;21(8):837–45. https://doi.org/10.1093/sleep/21.8.837. PMID: 9871946.
[17] Marks GA, Shaffery JP, Oksenberg A, et al. A functional role for REM sleep in brain maturation. Behav Brain Res 1995;69:1–11.
[18] Blumberg MS, Dooley JC, Sokoloff G. The developing brain revealed during sleep. In: Current opinion in physiology. vol. 15. Elsevier; 2020. p. 14–22. https://doi.org/10.1016/j.cophys.2019.11.002. Jun. 01.

[19] Guyer C, Huber R, Fontijn J, et al. Cycled light exposure reduces fussing and crying in very preterm infants. Pediatrics 2012;130(1):145–51. https://doi.org/10.1542/peds.2011-2671.
[20] Cao M, He Y, Dai Z, et al. Early development of functional network segregation revealed by connectomic analysis of the preterm human brain. Cerebr Cortex 2017;27(3):1949–63. https://doi.org/10.1093/cercor/bhw038.
[21] Ribas CG, Andreazza MG, Neves VC, et al. Effectiveness of hammock positioning in reducing pain and improving sleep-wakefulness state in preterm infants. Respir Care 2019;64(4):384–9.
[22] Karami S, Marofi M. A comparison between the effects of environmental and behavioral interventions on sleep cycle of preterm infants in NICU. J Kerman Univ Med Sci 2018;25(6):505–8.
[23] Loewy J, Stewart K, Dassler AM, et al. The effects of music therapy on vital signs, feeding, and sleep in premature infants. Pediatrics 2013;131(5): 902–18. https://doi.org/10.1542/peds.2012-1367.
[24] Volpe JJ. Dysmaturation of premature brain: importance, cellular mechanisms, and potential interventions. Pediatr Neurol 2019;95:42–66. https://doi.org/10.1016/j.pediatrneurol.2019.02.016.
[25] Werth J, Atallah L, Andriessen P, Long X, Zwartkruis-Pelgrim E, Aarts RM, et al. Unobtrusive sleep state measurements in preterm infants – a review. Sleep Med Rev 2017;32:109–22. https://doi.org/10.1016/j.smrv.2016.03.005.
[26] Munn Z, Peters MDJ, Stern C, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18:143. https://doi.org/10.1186/s12874-018-0611-x.
[27] Armstrong R, Hall BJ, Doyle J, et al. 'Scoping the scope' of a cochrane review. J Public Health 2011;33(1):147–50.
[28] Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7): 467–73. https://doi.org/10.7326/M18-0850.
[29] Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan—a web and mobile app for systematic reviews. 5 Syst Rev 2016;51(1):1–10. https://doi.org/10.1186/S13643-016-0384-4. Dec. 2016.
[30] Bik A, Sam C, de Groot E, et al. Protocol: a scoping review on behavioral sleep state classification methods. 2021.
[31] Thoman EB. Early development of sleeping behaviors in infants. In: Ellis NT, editor. " in Behavior and development in infancy: human and animal studies. London: Pennsylvania State; 1995.
[32] Parmelee Jr AH. Ontogeny of sleep patterns and associated periodicities in infants. Mod Probl Paediatr 1974;13:298–311.
[33] International Classification of Sleep Disorders: Diagnostic and coding manual. 2nd ed. Westchester: American Academy of Sleep Medicine; 2005.
[34] Holditch-Davis D, Blackburn ST, VandenBerg K. Newborn and infant neurobehavioral development. In: Kernner C, Lotte JW, editors. " in Comprehensive neonatal nursing: a physiologic perspective. 3rd ed. Philadelphia, PA: Saunders; 2003. p. 236–84.
[35] Korner AF, Thom VA, Forrest T. The neurobehavioral assessment for preterm infants. The Psychological Corporation; 2005.
[36] Gavin B, Theodore DW. Handbook of infant development. In: Brazelton TB, Nugent BM, Lester JK, editors. " in Neonatal behavioral assessment scale. 2nd ed. 1987. p. 780–817.
[37] Prechtl HFR. The neurological examination of the fullterm newborn infant. Philadelphia PA: Lippincott; 1977.
[38] Sumner G, Spietz A. NCAST: caregiver/parent-child interaction teaching manual. Seattle, WA: NCAST Publications, University of Washington, School of Nursing; 1998.
[39] Edraki M, Zendehzaban S, Beheshtipour N, et al. "Comparison of the effects of attachment training for mothers on the behavioral responses of premature infants: a randomized clinical trial," Iran. J. Neonatol. 2015;6(2):37–42 [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L604977473&from=export U2 - L604977473.
[40] Curzi-Dascalova L, Mirmiran M. Manual of methods for recording and analyzing sleep-wakefulness states in preterm and full-term infant. Paris: INSERM; 1996.
[41] Liaw JJ, Yang L, Lo C, et al. Caregiving and positioning effects on preterm infant states over 24 hours in a neonatal unit in Taiwan. Res Nurs Health 2012;35(2): 132–45.
[42] Liaw JJ, Yang L, Lee CM, et al. Effects of combined use of non-nutritive sucking, oral sucrose, and facilitated tucking on infant behavioural states across heel-stick procedures: a prospective, randomised controlled trial. Int J Nurs Stud 2013;50(7):883–94.
[43] Prechtl HFR. Behavioral states of the newborn infant (a review). Brain Res 1974;76(2):185–212. https://doi.org/10.1016/0006-8993(74)90454-5.
[44] Brazelton TB. Neonatal behavioral assessment scale. 1st ed. London, UK: Clinics in Developmental Medicine; 1973.
[45] Parmelee Jr AH, Stern E. In: Clemente CD, Purpura DP, Mayer FE, editors. "Development of states in infants," in sleep and the maturing nervous system. Cambridge, Massachusetts: Academic Press; 1972. p. 199–228.
[46] Thoman E. In: Ellis N, editor. "Early development of sleeping behavior in infants," in Aberrant development in infancy. Hillsdale, NJ: Erlbaum; 1975. p. 132–8.
[47] Stefanski M, Schulze K, Bateman D, et al. A scoring system for states of sleep and wakefulness in term and preterm infants. Pediatr Res 1984;18(1):58–62.
[48] Holditch-Davis D. The development of sleeping and waking states in high-risk preterm infants. Infant Behav Dev 1990;13(4):513–31. doi: 10.1016/0163-6383(90)90020-9 LK, https://vu.on.worldcat.org/oclc/4929202204.
[49] Emde RN, Koenig KL. Neonatal smiling and rapid eye movement states. J Am Acad Child Psychiatr 1969;8(1):57–67. https://doi.org/10.1016/S0002-7138(09)61566-1.
[50] Emde RN, Koenig KL. Neonatal smiling, frowning, and rapid eye movement states: II. Sleep-cycle study. J Am Acad Child Psychiatr Oct. 1969;8(4):637–56. https://doi.org/10.1016/S0002-7138(09)62248-2.
[51] Giganti F, Hayes MJ, Akilesh MR, et al. Yawning and behavioral states in premature infants. Dev Psychobiol 2002;41(3):289–96.
[52] Brandon DH, Holditch-Davis D, Winchester DM. Factors affecting early neurobehavioral and sleep outcomes in preterm infants. Infant Behav Dev 2005;28(2):206–19 [Online]. Available: https://www.embase.com/search/results?subaction=viewrecord&id=L40637381&from=export U2 - L40637381.
[53] Anderson GC, Behnke M, Gill NE, et al. Self-regulatory gavage to bottle feeding for preterm infants: effect on behavioral state, energy expenditure, and weight gain. In: Key aspects of recovery: improving nutrition, rest, and mobility; 1990. p. 83–97.
[54] Korner AF. State as a variable, as obstacle, and as mediator of stimulation in infant research. Merrill Palmer Q Behav Dev 1972;18(2):77–94.
[55] Holditch-Davis D, Scher M, Schwartz T, et al. Sleeping and waking state development in preterm infants. Early Hum Dev 2004;80(1):43–64.
[56] de Groot ER, Knoop MS, van den Hoogen A, et al. The value of cardiorespiratory parameters for sleep state classification in preterm infants: a systematic review. Sleep Med Rev 2021;58:101462. https://doi.org/10.1016/j.smrv.2021.101462.
[57] Eichenwald EC, Ungarelli RA, Stark AR. Hypercapnia increases expiratory braking in preterm infants. J Appl Physiol 1993;75(6):2665–70.
[58] Museux N, Cardot V, Bach V, et al. A reproducible means of assessing the metabolic heat status of preterm neonates. Med Phys 2008;35(1):89–100.
[59] Shepherd KL, Yiallourou SR, Odoi A, et al. When does prone sleeping improve cardiorespiratory status in preterm infants in the NICU? Sleep 2020;43(4): zsz256.
[60] Telliez F, Bach V, Dewasmes G, et al. Sleep modifications during cool acclimation in human neonates. Neurosci Lett 1998;245(1):25–8.
[61] Wolff PH. Observations on newborn infants. Psychosom Med 1959;21(2): 110–8.
[62] Hendricks JC, Kline LR. Differential activation within costal diaphragm during rapid-eye-movement sleep in cats. J Appl Physiol 1985;70(3):1194–200.
[63] Kuhn P, Zores C, Pebayle T, et al. Infants born very preterm react to variations of the acoustic environment in their incubator from a minimum signal-to-noise ratio threshold of 5–10 dBA. Pediatr Res 2012;71(4–1):386–92. https://doi.org/10.1038/pr.2011.76.
[64] Georgoulas A, Jones L, Laudiano-Dray MP, et al. Sleep-wake regulation in preterm and term infants. Sleep 2021;41(1). https://doi.org/10.1093/sleep/zsaa148. zsaa148.
[65] Grigg-Damberger M, Gozal D, Marcus CL, et al. The visual scoring of sleep and arousal in infants and children. J Clin Sleep Med 2007;3(2):201–40. https://doi.org/10.5664/jcsm.26819.