INTENS CRIT CARE NUR xxx (xxxx) xxx

Contents lists available at ScienceDirect

Intensive & Critical Care Nursing

journal homepage: www.sciencedirect.com/journal/intensive-and-critical-care-nursing

Review Article

Instruments to measure complexity of care based on nursing workload in intensive care units: A systematic review

Cristina Reguera-Carrasco *, Sergio Barrientos-Trigo

Department of Nursing, Faculty of Nursing, Physiotherapy, and Podiatry, Universidad de Sevilla, C/ Avenzoar, 6, 41009 Seville, Spain

A B S T R A C T

Objective: To establish an evidence-based recommendation on the use of validated scoring systems that measure nursing workload in relation to the complexity of care in adult Intensive Care Units.

Methods: A systematic review based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) was conducted (PROSPERO registration: CRD42021251272). We searched for validation studies until July 2023 using the bibliographic databases CINAHL, Scopus, Pubmed, WOS, Cochrane Database, SCIELO, Cuiden and Cuidatge. Reference selection and data extraction was performed by two independent reviewers. The assessment of risk of bias was performed using QUADAS-2 and the overall quality according to COSMIN and GRADE approach.

Results: We included 22 articles identifying 10 different scoring systems. Reliability, criterion validity and hypothesis testing were the most frequently measurement properties reported. The NAS was the only tool to demonstrate a Class A recommendation (the best performing instrument).

Conclusions: NAS is the best currently available scoring system to assess complexity of care from nursing workload in ICU. However, it barely met the criteria for a class A recommendation. Future efforts should be made to develop, evaluate, and implement new systems based on innovative approaches such as intensity or complexity of care.

Implications for Clinical Practice: The results facilitate decision making as it establishes a ranking of which instruments are recommended, promising or not recommended to measure the nursing workload in the intensive care units.

Introduction

Complexity of care (CC) is a concept widely used in current scientific literature that identifies situations that cause uncertainty and unpredictability within healthcare systems. Complexity factors have been defined as those related to patients, organization, and operators (Guarinoni et al., 2015; Guarinoni et al., 2014). The patient's components are related to their general condition and their clinical situation (Adamuz et al., 2018; Huber et al., 2021). At the organizational level with the healthcare practice environment (Busnel et al., 2022; Guarinoni et al., 2015) and at the operator level it is related to their skills, professional experience and, especially, to workloads (Hoeve et al., 2018).

Workload is a factor variable to evaluate the complexity of care due to its relationship with complexity factors (Adamuz et al., 2018; Griffiths et al., 2020). In fact, individual factors such as the sex of the patient; organizational factors such as the number of patients under care; and operator factors such as the sex of nurse, have demonstrated its relationship with increased workload (Hoogendoorn et al., 2021a,b; Moghadam et al., 2021).

Intensive Care Units (ICU) are units with a greater workload due to the patient's condition, the specificity and specialization of care (Silva et al., 2021). In the last three decades, studies that analyze the impact of nursing workloads on health outcomes of ICU patients have increased significantly (da Palma Afonso, 2018; Azevedo et al., 2022; Carrara et al., 2016; Queijo et al., 2013). Workload is associated with an increase between 1.5 % and 14 % in pressure injuries (Porcel-Gálvez et al., 2022), between 4 and 14 % in mortality and between 4 and 11 % in failure to rescue (Labelle et al., 2019).

In recent times, studies have proliferated whose objective was to validate scoring systems for measuring nursing workload. The proposed tools have been developed based on direct or indirect methods. The first group includes those that measure the time spent caring for the patient, such as the Time Oriented Score System (TOSS) or the Nursing Activities Score (NAS) (Hoogendoorn et al., 2020). The second group includes those that use other variables such as severity or activities performed (Valls-Matarín et al., 2015). Among them we could find, the Simplified Therapeutic Intervention Scoring System or the Nine Equivalents Manpower Score (Rollán et al., 2011).

Until now, reviews have been found in the literature on existing scoring systems for quantifying nursing work in the ICU (Rivera et al.,

* Corresponding author.
E-mail address: creguera@us.es (C. Reguera-Carrasco).

https://doi.org/10.1016/j.iccn.2024.103672
Received 1 December 2023; Received in revised form 22 February 2024; Accepted 2 March 2024

Please cite this article as: Cristina Reguera-Carrasco, Sergio Barrientos-Trigo, *INTENS CRIT CARE NUR*, https://doi.org/10.1016/j.iccn.2024.103672

2022; Rollán et al., 2011; Subirana & Solà, 2006). However, none of them have evaluated the quality of the validation studies, so recommendations for use based on the quality of their psychometric properties are unknown.

Therefore, the aim of our study is to establish an evidence-based recommendation regarding the use of nursing workload scoring systems in relation to the complexity of care in adult Intensive Care Units.

Methods

A systematic review based on PRISMA statement and COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) was used (Prinsen et al., 2018; Page et al., 2021). The protocol was registered in the PROSPERO database (registration number: CRD42021251272). Meta-analysis was not feasible because of the wide variation in how both staffing and outcomes were measured.

Search strategy

Fourth steps were followed in the search strategy. First, we conducted primary searches in PubMed using both MeSH terms and free terms. An example of our PUBMED search strategy is included in the e-component 1. The identified search strategy was confirmed by our research group. Second, we conducted a primary search in CINAHL, Scopus, Pubmed, WOS, Cochrane Database, SCIELO, Cuiden y Cuidatge. In addition, the search strategy was adapted for other databases. Third, all the bibliographic references of the articles evaluated in full text were analyzed. Fourth, a second search was performed in CINAHL using the terms "workload", "tool" and "validation studies" related to the names of the scales found in the first search.

Inclusion and exclusion criteria

The inclusion criteria were as follows: studies that aimed to develop or to validate nursing workload scoring systems; studies with information enough to evaluate one or more psychometric properties; studies that targeted patients over 18 years of age who were admitted in hospital critical care units, whether intensive or semi-intensive care units; studies published in English, Spanish, Italian, Catalan and Portuguese with no lower date limit were included. We searched all literature up till July 1st, 2023. The exclusion criteria included the following: unavailability of the full text, studies conducted in a mixture of units, like NICUs, PICUs and adult ICUs if their analysis was not reported separately.

Study screening and selection

Citations were managed by two researchers independently using RAYYAN systematic review software. They screened references first by reading titles and abstracts and then by reviewing full texts. The disagreements throughout the study selection process were resolved by consensus.

Data extraction

Two researchers independently extracted data from the included studies. Tables 1 and 2 included information about: the name of the scoring systems, its author, year of publication, country, location, target population, outcome measure, content validity, construct and criterion validity and internal consistency.

Synthesis and integration

No statistical analysis has been carried out in this systematic review. A narrative synthesis of the main characteristics of the scoring systems' properties has been made, as well as which them are recommended or

Table 1
Scoring systems.

Scoring Systems (year 1st publication)	Range of scores /Frequency	Scoring Methods
VACTE (2007) Scale of Workload and Nursing Times (Braña et al., 2007)	230–895/Per 24 h	- Description of nursing contribution in a Semi-intensive Care Unit in 13 categories of nursing procedures - Total points categorized to estimated nursing time
NCR (1992) The Nursing Care Recording System (Hjortso et al., 1992)	0–30 /Per shift	- 10 indicators of nursing care procedures - 1–3 points per indicator - Total score categorized to estimated nursing time into three categories: A (stable patients), B (stable patients due to specific treatment) or C (unstable patients)
NCR-11 (2004) Modified Nursing Care Recording System (Walther et al., 2004)	0–30/Per shift	Updated version of the NCR - Description of nursing contribution to 11 categories of nursing care procedures - 1–3 points per category - Total points categorized to estimated nursing time
TISS (1974) Therapeutic Intervention Scoring System (Cullen et al., 1974)	Per 24 h	- 57 items of therapeutic intervention - 1–4 points per intervention - Total score categorized into four levels according to intensity of involvement
TISS-28 (1996) Simplified Therapeutic Intervention Scoring System (Miranda et al., 1996)	0–79/Per 24 h	Simplified version of the TISS-76 with 28 therapeutic medical interventions - 1–4 points per intervention. - Total score categorized into four levels of care
NEMS (1997) Nine Equivalents of nursing Manpower use Score (Miranda et al., 1997)	0–63/Per shift or per 24 h	Simplified version of the TISS-28 with 9 activities - 1–8 points per activity - Total score categorized into four levels of care
EVECTE (2003) Nursing Workload Assessment Scale (Padrón et al., 2003)	4–50/Per shift	- 15 categories related with medical and nursing procedures and nurse patient-ratio. - 4-grade nursing workload: ideal, adequate, large, excessive and enormous
PINI (1989) The Patient Intensity forNursing Index (Prescott et al., 1989)	Per shift	- 12 dimensions of nursing care - Level I–V per dimension, representing the complexity of nursing care
CNIS (2003) The Comprehensive Nursing Intervention Score (Yamase, 2003)	Per 24 h	- List of 73 nursing interventions - 4-grade workload score in 6 aspects per intervention: nursing time needed, number of nurses, muscular extension, mental stress, skill, job intensity
NAS (2003) The Nursing Activities Score (Miranda et al., 2003)	0–177 %/Per shift or per 24 h	- 23 nursing activities - A score 1.2–32 points per nursing activity - Points representing the required nursing time per activity

Table 2
Scoring system psychometric property results.

Scoring Systems	*Reference/Country*	*Sample size*	*Content Validity*		*Construct and Criterion Validity*	*Internal Consistency*
VACTE	Braña-Marcos (2007)/ Spain (Spanish)	N= 121 adults in Intermediate care unit	13 ítems (71 sub-ítems) *Items:* 1.Monitoring 2. Basics care 3. Respiratory support 4. Nutrition 5.Movilization 6. Psicologist support	7. Wound healings 8. Insulations 9. Medication 10. Routine techniques 11. Invasive techniques 12. Sample collection 13. Others.	*Criterion Validity* NEMS: r = 0.49; p < 0.001 APACHE: r = 0.43; p < 0.001	–
NCR	Hjortso (1992) / Denmark (English)	N= 393 adults in ICU	10 indicators; 3 points Likert Scale *Indicators:* 1. Frequency of monitoring 2. Central nervous system 3. Respiration 4.Cardiovascular 5.Gastrointestinal	6. Renal 7. Infusions and injections 8. Patient care 9. Other monitoring and treatment assist 10. Hematology	*Criterion Validity* Therapeutic Intervention Scoring system (TISS): r=0.60; P<0.05.	-
NCR-11	Walther (2004)/ Sweden (English)	N= 6116 adults in ICU	11 categories; 3 points Likert Scale *Categories:* 1. Respiratory system 2. Circulatory system 3. Renal system 4. Central nervous system 5. Monitoring 6. Infusions/Injections 7. Samples	8. Nursing care 9. Other monitoring or treatment 10. Dressings, drain/ Ostomies 11. Care of relatives	*Convergent Validity* NCR11 vs. TISS Unit A: ICU-day/24h/All stay ICU <0.001 r = 0.67 / 0.61 / 0.99 ICU B: ICU-day/ 24h/All stay ICU <0.001 r = 0.58 / 0.55 / 0.98 NCR11 vs. NEMS Unit A: ICU-day/24h/All stay ICU <0.001 r = 0.44 / 0.44 / 0.99 Unit C: ICU-day/24h/All stay ICU <0.001 r = 0.50 / 0.49 / 0.99	*Inter-observer reliability* CV total = 7.4% Caso 1: CV 10.4 Caso 2: CV = 5.9 Caso 3: CV = 5.9
TISS	Cullen (1974) / EEUU, (English)	Aproximately 850 patient/day.	57 items		-	-
TISS-28	Wang (2017)/ China, (English)	N= 133 adults in ICU (medical-surgical)	28 ítems (1-8 points) *Categories:* 1. Basic Activities 2. Ventilatory Support 3. Cardiovascular Support 4. Renal Support 5. Neurologic Support 6. Metabolic Interventions 7. Specific Interventions	*Expert Validity* S-CVI/UA 0.82 S-CVI/Ave 0.94	*Criterion Validity* Therapeutic Intervention Scoring System 76 items (TISS-76) r = 0.764; P = <0.001 *Construct Validity* Workload first day vs last day: 30.76 ± 6.86 vs 24.67 ± 5.48; P < 0.001 *Convergent Validity* APACHEII r = 0.432; P < 0.001	*Inter-observer reliability* ICC 0.959; P < 0.001
	Miranda (1996) / Norway, (English)	N= 906 adults in ICU (general and medical-surgical)	28 ítems *Categories:* 1. Basic Activities 2. Ventilatory Support 3. Cardiovascular Support	4. Renal Support 5. Neurologic Support 6. Metabolic Interventions 7. Specific Interventions	EFA: 11 factors explain 61,7 % of variance. *Criterion Validity* Therapeutic Intervention Scoring System 76 items (TISS-76) r = 0.93 P = 0.001	-

(*continued on next page*)

Table 2 (*continued*)

Scoring Systems	*Reference/Country*	*Sample size*	*Content Validity*		*Construct and Criterion Validity*	*Internal Consistency*
	Castillo-Lorente (1999) / Spain, (English)	N= 8838 ICU patients	28 ítems	*Criterion Validity*	-	
					Therapeutic Intervention Scoring System 76 items (TISS-76) r = 0.85; p < 0.0001)	
	Moreno (1997) / Portugal, (English)	N= 1094 adults in ICU (medical-surgical)	28 ítems	*Convergent Validity*	*Inter-observer reliability*	
				TISS-76: R2 = 0.72 P > 0.001 (TISS-28 explains 72% of variance TISS-76).	TISS - 28 ICC = 0.93; IC (0.83 - 0.98) TISS-76 ICC = 0.95; IC (0.86 to 0.98)	
	Kwok (2005) / China, (English)	N= 156 adults in ICU (general)	28 ítems (1-8 points)	*Criterion Validity*	*Inter-observer reliability*	
			Expert Validity	TISS 76: r = 0.78; P = 0.001	ICC 1.00	
			CVI = 0.86	SAPS II: r = 0.68; P = 0.001 *Construct Validity* ICU patient vs. Rehabilitation: 28.39 vs. 2.84 (r = 0.68, P = .001).		
NEMS	Miranda (1997) / Norway, (English)	N= 996 adults in ICU (general and medical-surgical)	9 ítems	*Convergent Validity*	Total scale:	
			1. Basic monitoring	TISS-28: R2 = 0.7559 p < 0.001	ICC 0.92	
			2. Intravenous medication	(NEMS explains 75% of variance TISS-28).	k= 0.734 - 0.919 %	
			3. Mechanical ventilatory support	SAPII: r = 0.55 < 0.001 (NEMS explains	C = 90% - 99.4%	
			4. Supplementary ventilatory care	30.4% of variance SAP II).		
			5. Single vasoactive medication			
			6. Multiple vasoactive medication			
			7. Dialysis techniques			
			8. Specific interventions in the ICU			
			9. Specifics intervention outside the ICU			
EVECTE	Sánchez (2003) / Cuba, (Spanish)	N= 100 adults in ICU	15 ítems		-	-
			3 points Likert Scale			
			Categories			
			1. Admitted patients	8. Vital signs		
			2. Nursing staff	9. Monitoring		
			3. Hydration activities	10. Drainage		
			4. Intravenous medication	11. Hydric balance		
			5. Intramuscular or Subcutaneous	12. Level or awareness		
			6. Oral care	12. Cures		
			7. Digestive activities	13. Hygiene		
				14. Mechanical ventilatory support		
PINI	Prescott (1991) / EEUU, (English)	N= 6445 inpatients	10 ítems; 5 points Likert Scale	EFA: 3 factors		Total scale:
		Items		*Convergent Validity*		$\alpha = 0.85$
		1. Severity of illness	6. Emotional status	1. PINI vs medical diagnosis: r = 0.33*	4. PINI vs stay:	*Inter-observer reliability*
		2. Physiological status	7. Knowledge deficit	2. PINI vs. consults: r = 0.17**	r = 0.31**	Kw = 0.62
		3. Activities of daily living	8. Task/procedure complexity	3. PINI vs. severity of lllness: = r = 0.44**	5. PINI vs. disposition: F (3,116) = 14.36**	
		4. Mobility	9. Complexity of clinical judgments		6. PINI vs. PSC	
		5. Potential for injury	10. Hours of care		- MEDICUS r = 0.70** - GRASP r = 0.54** - San Joaquin System r = 0.55**	

(*continued on next page*)

Table 2 (*continued*)

Scoring Systems	*Reference/Country*	*Sample size*	*Content Validity*		*Construct and Criterion Validity*	*Internal Consistency*
	Prescott (1989) / EEUU, (English)	N= 657 inpatients	12 ítems; 5 points Likert Scale Ítems 1. Severity of illness 2.Complications/Other conditions 3. Response to treatment 4. Activities of daily living 5. Mobility 6. Physiological status	- 7. Emotional sneed 8. Teaching 9. Sensory/ Communicative/ Cognitive Limitation 10. Task/procedure complexity 11. Complexity of clinical judgments 12. Hours of nursing care	*Inter-observer reliability* kw = 0.45 - 0.86 kmax = 0.80 - 0.92	
CNIS	Yamase (2003) / Japan, (English)	N= 107 adults in ICU	73 items (8 categories); 3- points Likert Scale *Categories:* 1. Monitoring 2. Transfusion of blood/fluids 3. Injections 4. Respiratory management	*Criterion Validity* NEMS: r = 0.55; p < 0.001 5. Assisted circulation 6. Drainage tube management 7. Special therapy 8. Basic nursing care	*Inter-observer reliability* K = 0.65; p > 0. 01	
NAS	Miranda (2003) / Netherland, (English)	N= 2041 ICU patients	23 items (4 categories) *Categories:* 1. Nursing activities at patient level 2. Nursing activities not relating to a patient 3. Personal activities for the nurse 4. Other categories	*Criterion Validity* TISS-28: r= 0.56 (p < .001).	-	
	Lachance (2020) / Canada (English)	N= 155 adults in ICU (heart and/or lung disease)	23 ítems (7 dimensions) *Dimensions* 1. General activities 2. Respiratory 3. Cardiovascular	 4. Renal 5. Neurological 6. Metabolic 7. Specific interventions	*Convergent Validity* NAS at admission in CCU and PICU; p < 0.001 - APACHE: r = 0.091 ; r = 0.379 - SAPS 3: r = 0.181 ; r = 0.331 - LOS: r = 0.220; r = 0.207 - NAO: r = 0.272; r = 0.043	-
	Lachance (2018) / Canada (English)	N= 15 ICU patients	23 ítems (7 dimensions) *Dimensions* 1. General activities 2. Respiratory 3. Cardiovascular 4. Renal	*Expert Validity* CVI = 0.2 - 1 5. Neurological 6. Metabolic 7. Specific interventions	-	*Inter-observer reliability* ICC 0.90; P < 0.001
	Macedo (2016) / Brazil, (English)	N= 67 adults in ICU	23 ítems (7 dimensions) *Dimensions:* 1. Basic activities 2. Ventilatory support 3. Cardiovascular support 4. Renal support	 5. Neurologic support 6. Metabolic support 7. Specific interventions	EFA: 7 factors explain 63.8% of variance.	Total scale: $\alpha = 0.71$ Factor 1: Basic activities: a = 0.52 Factor 2: Cardiovascular support: a = 0.05 Factor 3: Renal support: a= 0.04 Factor 4: Specific interventions: a = 0.24 Factor 5: Ventilatory support: a = 0.60 Factor 6: Metabolic support: a = 0.76
	Sánchez-Sánchez (2015) / Spain, (Spanish)	N= 1046 adults in ICU (general and burn unit)	*Sub-items:*	8. Administrative tasks	KMO = 0.589	Total scale: $\alpha = 0.373$

(*continued on next page*)

Table 2 (*continued*)

Scoring Systems	*Reference/Country*	*Sample size*	*Content Validity*		*Construct and Criterion Validity*	*Internal Consistency*
			1. Monitoring	9. Ventilatory support	EFA: 7 factors explain 70% of variance	*Intra-observer reliability*
			2. Laboratory tests	10. Cardiovascular support.		ICC = 0.837; IC (0,466-0,950; p = 0,002).
			3. Medication therapy	11. Renal support	*Convergent Validity*	*Inter-observer reliability*
			4. Hygiene	12. Neurologic support	NEMS: (r = 0.719; R2 = 0,5126)	ICC = 0.662; IC (0,033-0,882; p = 0.001)
			5. Drainage	13. Metabolic support		K= 0.371; C = 44%
			6. Posture changes	14. Specific interventions		
			7. Family support			
	Arias-Rivera (2013)/ Spain, (Spanish)	n= 30 ICU patients	23 ítems		-	-
	Toffoletto (2018) / Chile, (Spanish)	N= 126 adults in ICU	7 categories (23 items)		KMO = 0.699	Total scale: KR = 0.886
			Expert consensus CVI: 0.6 - 1.0		Bartlett's sphericity test P < 0.0001	
			Lyn index: 80 - 100		Inter-factor Correlation: 0.003 - 0.872	
					EFA: 1 factor explains 12.8% of variance	
					Criterion Validity	
					TISS-28: r = 0.601; p < 0.0001	
	Queijo (2009) / Portugal, (English)	N= 100 adults in ICU	23 ítems		*Criterion Validity*	Total scale: α = 0.36
					Therapeutic Intervention Scoring System (TISS-28): r = 0,67; P < 0.001	Inter-observer reliability
						K = 0.99
					Convergent Validity	
					SAP II: (R2= 99,8%; p < 0.001)	

EFA: Exploratory factor analysis; CVI: Content Validity Index; KMO: Kaiser-Meyer.Olkin; CV: Coefficient of variation: %C: % Concordance; KR: Kuder-Richardson; ICC: Intraclass Correlation Coefficient; K = kappa; kw = Kappa weighted; r = Pearson correlation coefficient; ICU: Intensive Care Unit.
*(p < 0.001); ** (p < 0.0001).

not, depending on the quality of their measured psychometric characteristics.

Assessment of methodological quality

Two researchers independently assessed the RoB of each study and evaluating the measurement quality of each measurement property according to the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) and the COSMIN risk of bias checklist.

The QUADAS-2 analyzes the RoB and applicability of the studies based on four domains: patient selection, index test, reference standard, flow and timing (Whiting et al., 2011).

The COSMIN Risk of Bias checklist was developed for assessing the methodological quality of single studies included in systematic reviews of scoring systems. The checklist contains standards referring to design requirements and preferred statistical methods of studies on measurement properties. The COSMIN Risk of Bias checklist contains 10 domains and 116 items. It evaluates different psychometric properties: content validity, structural validity, internal consistency, reliability, measurement error, hypothesis testing, cross-cultural validity, criterion validity and responsiveness. The methodological quality of each study will be evaluated as "very good", "adequate", "doubtful", "inadequate", and "not applicable". The lowest rating of any item was used to determine the overall rating of each measurement property's quality. Finally, the measurement properties were rated as "sufficient (+)", "insufficient (−)", or "indeterminate (?)" based on the criteria for good measurement properties of COSMIN (Mokkink et al., 2018).

Summarizing the quality of the psychometric properties

Two researchers independently summarized the quality of psychometric properties for each scoring system and came to an overall conclusion of its quality according to the COSMIN criteria. If the results per study were all sufficient (or all insufficient), the overall rating was also sufficient (or insufficient). If the results were inconsistent, to rate the qualitatively summarized results as sufficient (or insufficient), in principle 75 % of the results should met the criteria (Mokkink et al., 2018).

Grading the quality of evidence

Two researchers independently assessed the certainty of the evidence according to the modified Grading of Recommendations, Assessment, Development, and Evaluation system (GRADE) (Schünemann et al., 2013). Each psychometric property was graded using four factors (risk of bias, inconsistency, indirectness and imprecision) into four levels: high, moderate, low, or very low evidence. Finally, the included scoring systems were recommended into three categories based on the methodological quality and results of each psychometric property. If evidence showed that the scoring system had sufficient content validity (any level) and at least low-quality evidence for sufficient internal consistency, it was categorized as "strongly recommended". If the scoring system had high quality evidence for an insufficient measurement property, they were categorized as "not recommended". Other situations were categorized as "weakly recommended" (Mokkink et al., 2019; Prinsen et al., 2018).

Results

Literature search

The primary electronic search yielded 3557 articles. After removing duplicates, completing preliminary screening and reviewing full texts, we included 9 articles. The secondary search yielded a total of 13 articles. Finally, 22 articles and 10 scoring systems were selected. Fig. 1 shows a PRISMA flow chart with the results of the literature search strategy (Page et al., 2021).

Study and scoring systems description

The main characteristics of the studies and tools are shown in Table 1 and 2. Among the 22 articles included, 8 were original tools, 6 were extensions, 2 were adaptations and 6 were translations of the 8 originals. The eight original scoring systems are the Nursing Activity Score (NAS) (Miranda et al., 2003), the Comprehensive Nursing Intervention Score (CNIS) (Yamase, 2003), the Patient Intensity for Nursing Index (PINI) (Prescott et al., 1989), the Nine Equivalents of nursing Manpower use Score (NEMS) (Miranda et al., 1997), The Nursing Workload Assessment Scale (EVECTE) (Padrón et al., 2003), the Therapeutic Intervention Scoring System (TISS) (Cullen et al., 1974), the Nursing Care Recording (NCR) (Hjortsø et al., 1992), the Scale of Workload and Nursing Times (VACTE) (Braña et al., 2007).

The NAS was translated into Portuguese (Ferreira & Grillo, 2009; Macedo et al., 2016), Chilean (Toffoletto et al., 2018), Spanish (Sánchez-Sánchez et al., 2015) and French-Canadian (Lachance et al., 2018, 2020). The TISS (Cullen et al., 1974) was adapted into TISS-28 (Castillo-Lorente et al., 2000; Kwok et al., 2005; Miranda et al., 1996; Moreno & Morais, 1997) and translated into Chinese (Wang et al., 2018). The NCR was adapted into NCR-11 (Walther et al., 2004).

Related to the nursing workload outcome measured, two scoring systems assessed nursing time activities (NAS, VACTE), four measured severity of patient illness (TISS, TISS-28, EVECTE, NEMS), one measured intensity of care (PINI), two measured nursing care needs (NCR, NCR-11), and one use a mix method (CNIS).

Quality assessment

Risk of bias

The risk of bias (RoB) and applicability in the four domains, are shown in e-component-2. Worse scores were obtained for risk of bias than for applicability. Risk of bias was well controlled in 44 % of the domains, compared to 89 % for applicability. Related to RoB, the domain with the worst control was the flow domain. Also, in 68 % of the articles, the data necessary to measure the methodological quality are not reported.

In the patient selection domain, 4 studies at high RoB were considered. Three of them because they did not include all eligible consecutive participants or selected a convenience sample (Lachance et al., 2020; Prescott et al., 1991; Toffoletto et al., 2018) and one of them because it used a retrospective analysis (Braña et al., 2007). Also, EVECTE was considered with high applicability concerns because the scale was interpreted and designed by non-ICU professionals (Padrón et al., 2003).

Regarding to index test and reference test domain, we obtained unclear RoB in eleven studies and thirteen studies respectively, because they did not report enough data to determine whether the tests were applied independently. According to applicability, we considered high applicability concerns in one study (Braña et al., 2007) and unclear applicability concerns in three studies because they did not report sufficient data on the application of the tools (Padrón et al., 2003; Prescott

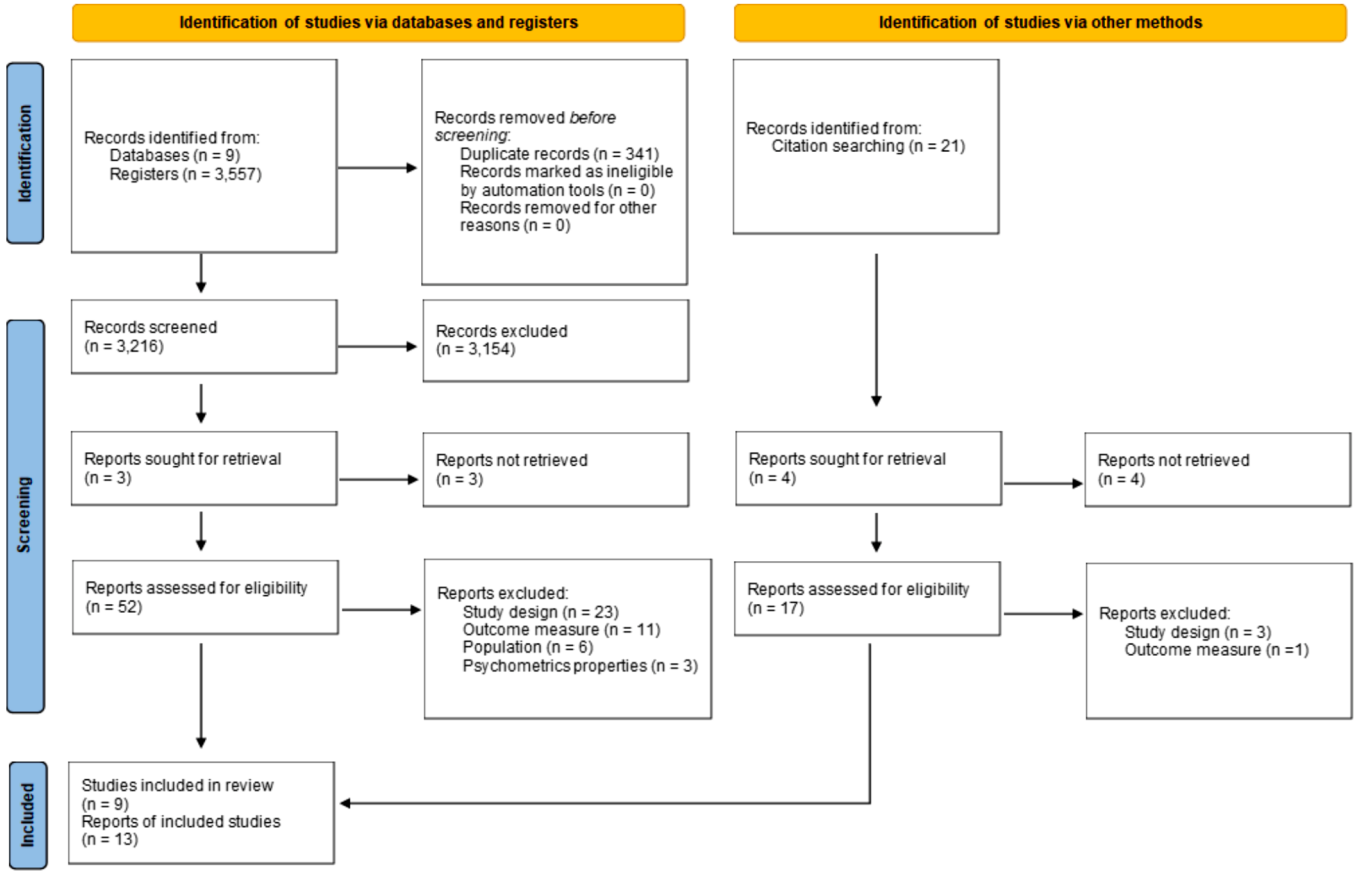

Fig. 1. Flow Diagram.

et al., 1991; Toffoletto et al., 2018).

Finally, in the Flow and timing domain, eight studies were considered with high RoB because the interval used for test administration was not adequate or not all patients were included (Hjortsø et al., 1992; Lachance et al., 2018; Miranda et al., 1997; Moreno & Morais, 1997; Prescott et al., 1991; Sánchez-Sánchez et al., 2015; Toffoletto et al., 2018).

Methodological quality assessment of measurement properties

The quality of psychometric properties retrieved from the included studies are shown in e-component-3. Of the 22 studies, only one article explored almost five psychometric properties (Ferreira & Grillo, 2009). The most reported psychometric properties were reliability, criterion validity and hypothesis testing. Limited information was retrieved on structural validity, internal consistency and cross-cultural validity. No data were identified for measurement error and responsiveness.

Regarding to content validity, from the perspective of professionals, mostly used an appropriate method for assessing the relevance of each item for the construct of interest, and the comprehensiveness of the scoring system. In related to internal consistency, is reported in 4 studies, (Ferreira & Grillo, 2009; Macedo et al., 2016; Sánchez-Sánchez et al., 2015; Toffoletto et al., 2018), and all are scored doubtful except (Macedo et al., 2016). Regarding reliability, three studies were of questionable methodological quality because they did not report the kappa index (Walther et al., 2004) or an appropriate time interval (Moreno & Morais, 1997) or because weighted Kappa was < 0.70 (Prescott et al., 1989).

Considering criterion validity, five studies reported a good or very good methodological quality (Castillo-Lorente et al., 2000; Hjortsø et al., 1992; Kwok et al., 2005; Miranda et al., 1996; Miranda et al., 2003; Toffoletto et al., 2018; Wang et al., 2018; Yamase, 2003). On the other hand, four studies achieved a doubtful methodological quality due to the scoring system and gold standard were not administered at the same time or because the way missing items will be handled was not clearly described (Braña et al., 2007; Ferreira & Grillo, 2009; Miranda et al., 1997; Prescott et al., 1991). Hypothesis testing was rated as sufficient (+) because they reported the hypothesis question (Braña et al., 2007; Ferreira, et al., 2009; Hjortso et al., 1992; Kwok et al., 2005; Lachance et al., 2020; Miranda et al., 1997; Prescott et al., 1991; Sánchez-Sánchez et al., 2015; Walther et al., 2004; Wang et al., 2018).

Finally, cross-cultural validity, was performed in 6 studies, and all of them obtained inadequate or doubtful quality and were rated as indeterminate (?) because multiple group factor analysis or DIF analysis did not performed (Arias-Rivera et al., 2013; Ferreira & Grillo, 2009; Lachance et al., 2018; Macedo et al., 2016; Toffoletto et al., 2018; Wang et al., 2018).

Overall quality of measurement properties

The final ratings for the measurement properties and the quality of evidence for the ten scoring systems are presented in Table 3. None of the instruments were evaluated for all seven measurement properties, considering that measurement error and responsiveness were not reported for any instrument. Only two of them, the NAS and TISS-28, were evaluated on the five measurement properties. In terms of quality of evidence, the NAS, TISS-28 and NEMS were considered high quality for reliability. For criterion validity, the NAS, NCR, TISS-28 and CNIS were regarded as high quality and about content validity, the TISS-28 and CNIS were considered high quality. The remaining properties like hypothesis testing, cross- cultural validity, structural validity and internal consistency were of low to moderate quality mostly.

Certainty of evidence and recommendations

According to the COSMIN categorization criteria, we have categorized the included scoring systems into three categories, presented in Table 3. Category A included those with evidence for sufficient content validity (any level) and at least low quality evidence for sufficient internal consistency; Category B, included some scoring systems not categorized in A or C; Category C, included some tools with high quality evidence for an insufficient measurement property. Regarding this criterion, the NAS were strongly recommended (category A). The NCR, NCR-11, TISS-28, NEMS, CNIS and PINI were weakly recommended (category B) because were considered promising, but still need further validation. Finally, the VACTE, EVECTE and TISS were not recommended (category C) (Mokkink et al., 2019).

Discussion

This study aimed to establish an evidence-based recommendation on the use of validated scoring systems that measure nursing workload in relation to the complexity of care in adult ICU. Until now, this is the first review to use COSMIN checklist for assessing evidence-based recommendations regarding the use of nursing workload scoring systems in relation to the complexity of care in ICUs. We identified 22 studies and 10 scoring systems.

The nursing workload in the intensive care units has become particularly relevant in recent years due to the COVID-19 pandemic (Bruyneel et al., 2022). During this period, ICU-nurses were confronted with new challenging working scenarios that altered the complexity of care in all its spheres; related to the patient, the operator, and the organization. Consequently, an increase in nursing workload during this period has been demonstrated, with an increase of the NAS per nurse of 98 % (Hoogendoorn et al., 2021a,b). However, despite the increasing workloads in recent years and the large volume of publications, data on nursing staffing methods are very limited (Griffiths et al., 2020). Most of

Table 3
Overall quality of evidence per each measurement property.

Scoring Systems		N studies	Content Validity	Structural Validity	Internal Consistency	Cross-cultural Validity	Reliability	Criterion Validity	Hypothesis Testing
(A)	NAS	8	Moderate	Moderate	Low	Low	High	High	Moderate
(B)	NCR	1	Moderate	NA	NA	NA	NA	High	Moderate
	NCR-11	1	Moderate	NA	NA	NA	Moderate	NA	Moderate
	TISS-28	5	High	Moderate	NA	Low	High	High	Moderate
	NEMS	1	Moderate	NA	NA	NA	High	Moderate	High
	CNIS	1	High	NA	NA	NA	NA	High	NA
	PINI	2	Moderate	Low	NA	NA	Moderate	Low	Moderate
(C)	VACTE	1	Low	NA	NA	NA	NA	Low	Low
	TISS	1	Moderate	NA	NA	NA	NA	NA	NA
	EVECTE	1	Low	NA	NA	NA	NA	NA	NA

Evidence quality based on a modified GRADE approach (Mokkink et al., 2018).
NA: not applicable; NAS: Nursing Activity Score; NCR:The Nursing Care Recording System; TISS:Simplified Therapeutic Intervention Scoring System; NEMS: Nine Equivalents of nursing Manpower use Score; CNIS: The Comprehensive Nursing Intervention Score; PINI: The Patient Intensity for Nursing Index; VACTE: Scale of Workload and Nursing Times; EVECTE: Nursing Workload Assessment Scale.
Recommendation: A: Strongly Recommended; B: Weak recommended; C: Not recommended.

the tools are originally developed for other applications or by national organizations that do not extend them beyond their country of origin. In fact, all instruments except the NAS, TISS-28, and PINI were only assessed in a single study, in line with the findings of Greaves's systematic review. They concluded that none of the selected measurement instruments can be recommended because of the almost total absence of reports of daily use (Greaves et al., 2018).

In relation to psychometric properties, none of the scoring systems was evaluated based on the seven measurement properties of the COSMIN (Mokkink et al., 2018). Measurement error and responsiveness were not assessed in any of the included studies and internal consistency was only assessed in the NAS scale. In general, each selected validation study reported few psychometric properties and only the Portuguese version of NAS (Ferreira & Grillo, 2009) evaluated at least 5 psychometric properties in the same validation. This may be because some of these scoring systems were developed several years ago, and COSMIN were not available at that time for these authors to access. However, we found recent validations that have not used any specific instrument validation methodology (Lachance et al., 2018, 2020; Macedo et al., 2016; Toffoletto et al., 2018; Wang et al., 2018).

The methodological quality for each psychometric property was questionable or inadequate for most of the instruments, except for reliability and criterion validity. In terms of content validity, this is due to the lack of evaluation regarding the relevance of the items for the professionals. This is consistent with the findings of a systematic review, where the Content Validity Index was only described in 1 of the selected tools, with a value of 0.85, which fell below the considered threshold index of 0.9 (Hoogendoorn et al., 2020).

Regarding internal consistency, no factor analysis was conducted, and internal consistency was not calculated for each dimension separately. Concerning hypothesis testing, the consequences of the hypotheses were either unclear or not previously formulated (although it was possible to deduce the expected outcomes). In the context of structural validity, limitations were observed due to an insufficient sample size, the absence of missing values, or unclear information on the performed model. Lastly, in terms of cross-cultural validity, limitations arose from not providing a clear description of the group variables or the statistical method used.

Based on the quality of the evidence and the COSMIN criteria, category B comprised the largest number of tools (Mokkink et al., 2018). For the six measures that met category B (NCR, NCR-11, TISS-28, PINI, CNIS and NEMS), the main reason was the lack of at least low-quality evidence of sufficient internal consistency and the poor psychometric properties reported. However, although it did not analyze internal consistency, the TISS-28 presents a considerable number of validations and a good quality in psychometric properties and RoB. Despite this, based on our assessment of the current evidence, TISS-28 concerned mainly medical interventions selected by physicians based on patient's therapeutic effort (Cullen et al., 1974). Quantifying nursing work based solely on therapeutic effort provides a distorted view of the nursing reality. Similarly, the results presented in Polish ICUs, confirm that these types of tools cannot be used as a gold standard for planning nursing staffing (Wysokiński et al., 2013). The rest of the tools included in category B have few validation articles. However, some tools such as PINI or CNIS present a promising approach based on patient needs and the intensity of their care. In line with the latest critical care models that support the measurement of workloads based on a holistic view of the patient and their care (Ellis & Dark, 2020).

Finally, the NAS was included in A category, because it is the only one that analyzes internal consistency. Furthermore, it is the scoring system with the highest number of validations and psychometric properties analyzed. In fact, the NAS is one of the most utilized nursing workload tools and has been studied for its reliability, translation and validity across international health systems (Bruyneel et al., 2018; Bruyneel et al., 2019; Greaves et al., 2018; Lucchini et al., 2014; Padilha et al., 2015; Rivera et al., 2022; Ross et al., 2023; Stafseth et al., 2018). However, this method of quantification does not consider all dimensions of the complexity of care (Guarinoni et al., 2015) or factors such as years of experience, level of nursing competencies, teaching of nursing students, or assistance to new clinical staff. In addition, Palese et al., (2016), reassessed the face and content validity of the NAS and showed that it is not fully adequate for measuring current nursing activities in ICUs and its weights have been considered not fully adequate for scoring average nursing time consumption (Palese et al., 2016).

Finally, based on the quality of evidence and COSMIN criteria, the NAS scale presents the best balance in terms of validation quality, cross-cultural adaptations, and the measurement approach used, based on nursing activities. However, to comprehend nursing workload, we need valid and reliable measures to define it. Therefore, in future research, the use of specific validation methodologies such as COSMIN should be considered for the development of these tools. Additionally, it would be interesting to reformulate the tools with a more current nursing perspective, incorporating concepts such as the complexity of care as a key element for measuring nursing workloads and predicting patient health outcomes (Güven et al., 2024). This would enhance healthcare management in the face of increased nursing workload, reducing the risk of nurse burnout without compromising the quality of care (Bruyneel et al., 2022).

Limitations

This study has some limitations that need to be mentioned. We only included studies published in English and Spanish in the databases selected. Therefore, we may have missed relevant scoring systems. In addition, the quality appraisal is highly dependent on the reporting completeness and clarity of the included studies. For several studies, we found that the information was often poorly reported, sometimes leading to subjective interpretation. Finally, another point to consider is that we only included validation studies.

Conclusion

This systematic review establishes an evidence-based recommendation on which tools should be used to measure nursing workload in ICU. During its conduct, a dearth in the published validation studies was observed. In addition, no specific methodology for validation of instruments has been used and the quality of measurement properties were mostly moderate or low, except for reliability and criterion validity. Consequently, only the NAS scale could be recommended according to COSMIN criteria. However, this scale does not consider some dimensions of nursing care complexity. Furthermore, more studies are needed to continue the validation of the six promising scoring systems included in category B.

Future efforts should be made to develop, evaluate, and implement new measurement systems for nursing workload based on innovative approaches such as the complexity of care.

Role of the Funding source

The first author benefits from a grant for training university professors (FPU21/05541) from the Ministry of Education, Culture and Sports of the Government of Spain.

Ethical statement

This is a systematic review. No ethical approval was required or sough to carry out this review.

CRediT authorship contribution statement

Cristina Reguera-Carrasco: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Sergio**

Barrientos-Trigo: Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.iccn.2024.103672.

References

Adamuz, J., González-Samartino, M., Jiménez-Martínez, E., Tapia-Pérez, M., López-Jiménez, M.M., Ruiz-Martínez, M.J., Rodríguez-Fernández, H., Delgado-Hito, P., Juvé-Udina, M.E., 2018. Care complexity individual factors associated with hospital readmission: a retrospective cohort study. J. Nursing Scholarship 50 (4), 411–421. https://doi.org/10.1111/JNU.12393.

Arias-Rivera, S., Sánchez-Sánchez, M.M., Fraile-Gamo, M.P., Patiño-Freire, S., Pinto-Rodríguez, V., Conde-Alonso, M.P., Ferguson, N.D., Frutos-Vivar, F., 2013. Adaptación transcultural al castellano del nursing activities score. Enfermeria Intensiva 24 (1), 12–22. https://doi.org/10.1016/j.enfi.2012.10.002.

Azevedo, A.V., Tonietto, T.A., Boniatti, M.M., 2022. Nursing workload on the day of discharge from the intensive care unit is associated with readmission. Intensive Crit. Care Nurs. 69, 103162 https://doi.org/10.1016/j.iccn.2021.103162.

Braña, B., del Campo Ugidos, R., Fernández, E., de la Villa, M., 2007. Proposal of a new assessment scale of workload and nursing times (VACTE ©). Enferm. Intensiva 18 (3), 115–140.

Bruyneel, A., Guerra, C., Tack, J., Droguet, M., Maes, J., Miranda, D.R., 2018. Traduction sémantique en français et implémentation du nursing activities score en belgiqueMed. Intensive Reanim., 27 (3). https://doi.org/10.3166/rea-2018-0029.

Bruyneel, A., Tack, J., Droguet, M., Maes, J., Wittebole, X., Miranda, D.R., Pierdomenico, L.D., 2019. Measuring the nursing workload in intensive care with the nursing activities score (NAS): a prospective study in 16 hospitals in Belgium. J. Crit. Care 54, 205–211. https://doi.org/10.1016/j.jcrc.2019.08.032.

Bruyneel, A., Lucchini, A., Hoogendoorn, M., 2022. Impact of COVID-19 on nursing workload as measured with the nursing activities score in intensive care. Intensive Crit. Care Nurs. 69, 103170 https://doi.org/10.1016/j.iccn.2021.103170.

Busnel, C., Vallet, F., Ashikali, E.M., Ludwig, C., 2022. Assessing multidimensional complexity in home care: congruencies and discrepancies between patients and nurses. BMC Nurs. 21 (1), 1–10. https://doi.org/10.1186/S12912-022-00942-X/TABLES/2.

Carrara, F.S.A., Zanei, S.S.V., Cremasco, M.F., Whitaker, I.Y., 2016. Outcomes and nursing workload related to obese patients in the intensive care unit. Intensive Crit. Care Nurs. 35, 45–51. https://doi.org/10.1016/j.iccn.2015.12.003.

Castillo-Lorente, E., Rivera-Fernandez, R., Rodriguez-Elvira, M., Vázquez-Mata, G., 2000. TISS 76 and TISS 28: correlation of two therapeutic activity indices on a spanish multicenter ICU database. Intensive Care Med 26, 57–61.

Cullen, D., Civetta, J., Briggs, B., Ferrara, L.C., 1974. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. Crit. Care Med. 2 (2), 57–60.

Da Palma Afonso, E., 2018. Nursing workload in the intensive care unit – can it be measured? Intensive Crit. Care Nurs. 48, 1–2. https://doi.org/10.1016/j.iccn.2018.06.005.

Ellis, W.M., Dark, T., 2020. Developing a tool to measure work intensity. Nurs. Manage. 51 (2), 10–12. https://doi.org/10.1097/01.NUMA.0000617036.73190.96.

Ferreira, A., Grillo, K., 2009. Nursing Activities Score (NAS): Cross-cultural adaptation and validation to Portuguese language. Rev. Esc. Enferm. USP 43 (Esp), 1018–1025.

Griffiths, P., Saville, C., Ball, J., Jones, J., Pattison, N., Monks, T., 2020. Nursing workload, nurse staffing methodologies and tools: a systematic scoping review and discussion. Int. J. Nurs. Stud. 103 https://doi.org/10.1016/J.IJNURSTU.2019.103487.

Guarinoni, M.G., Motta, P.C., Petrucci, C., Lancia, L., 2014. Complexity of care: a concept analysis. Ann Ig 26 (3), 226–236. https://doi.org/10.7416/AI.2014.1981.

Guarinoni, M., Petrucci, C., Lancia, L., Motta, P.C., 2015. The concept of care complexity: a qualitative study. jphr.2015.588 J. Public Health Res. 4 (3). https://doi.org/10.4081/JPHR.2015.588.

Güven, B., Topçu, S., Hamarat, E., Ödül Özkaya, B., Güreşci Zeydan, A., 2024. Nursing care complexity as a predictor of adverse events in patients transferred from ICU to hospital ward after general surgery. Intensive Crit. Care Nurs. 82, 103637 https://doi.org/10.1016/j.iccn.2024.103637.

Hjortsø, E., Buch, T., Ryding, J., Lundstrøm, K., Bartram, P., Dragsted, L., Qvist, J., 1992. The nursing care recording system. a preliminary study of a system for assessment of nursing care demands in the ICU. Acta Anaesthesiol. Scand. 36 (7), 610–614. https://doi.org/10.1111/j.1399-6576.1992.tb03530.x.

Hoogendoorn, M.E., Brinkman, S., Spijkstra, J.J., 2021a. The objective nursing workload and perceived nursing workload in intensive care units: analysis of association. Int. J. Nurs. Stud. 114, 103852 https://doi.org/10.1016/j.ijnurstu.2020.103852.

Hoeve, Y.T., Brouwer, J., Roodbol, P.F., Kunnen, S., 2018. The importance of contextual, relational and cognitive factors for novice nurses' emotional state and affective commitment to the profession. a multilevel study. J. Adv. Nurs. 74 (9), 2082–2093. https://doi.org/10.1111/JAN.13709.

Hoogendoorn, M.E., Brinkman, S., Bosman, R.J., Haringman, J., de Keizer, N.F., Spijkstra, J.J., 2021b. The impact of COVID-19 on nursing workload and planning of nursing staff on the intensive care: a prospective descriptive multicenter study. Int. J. Nurs. Stud. 121, 104005 https://doi.org/10.1016/j.ijnurstu.2021.104005.

Greaves, J., Goodall, D., Berry, A., Shrestha, S., Richardson, A., Pearson, P., 2018. Nursing workloads and activity in critical care: A review of the evidence. Intensive Crit. Care Nurs. 48, 10–20. https://doi.org/10.1016/j.iccn.2018.06.002.

Hoogendoorn, M.E., Margadant, C.C., Brinkman, S., Haringman, J.J., Spijkstra, J.J., de Keizer, 2020. Workload scoring systems in the Intensive Care and their ability to quantify the need for nursing time: A systematic literature review. Int. J. Nurs. Stud. 101, 103408. https://doi.org/10.1016/j.ijnurstu.2019.103408.

Huber, E., Kleinknecht-Dolf, M., Kugler, C., Spirig, R., 2021. Patient-related complexity of nursing care in acute care hospitals - an updated concept. Scand. J. Caring Sci. 35 (1), 178–195. https://doi.org/10.1111/SCS.12833.

Kwok, W.W.Y., Chau, J.P.C., Le Low, L.P., Thompson, D.R., 2005. The reliability and validity of the therapeutic activity index. J. Crit. Care 20 (3), 257–263. https://doi.org/10.1016/j.jcrc.2005.06.004.

Labelle, J.B., Audet, L.A., Farand, P., Rochefort, C.M., 2019. Are hospital nurse staffing practices associated with postoperative cardiac events and death? A systematic review. PLoS One 14 (10). https://doi.org/10.1371/JOURNAL.PONE.0223979.

Lachance, J., Douville, F., Machado, E., Dallaire, C., Oliveira, H., Houle, J., Gallani, M., 2018. Cultural adaptation of the nursing activities score to the french-Canadian. The Canadian Journal of Critical Care Nursing 29 (3), 32–38.

Lachance, J., Douville, F., Oliveira, E.M., Dallaire, C., Houle, J., Gallani, M.C., 2020. The Canadian journal of critical care nursing • Canadian Association of Critical Care Nurses. Canadian J. Critical Care Nursing 31 (3), 12–22.

Lucchini, A., De Felippis, C., Elli, S., Schifano, L., Rolla, F., Pegoraro, F., Fumagalli, R., 2014. Nursing activities score (NAS): 5 years of experience in the intensive care units of an Italian University hospital. Intensive Crit. Care Nurs. 30 (3), 152–158. https://doi.org/10.1016/j.iccn.2013.10.004.

Macedo, A., Mendes, C., Candeias, A., Sousa, M., Hoffmeister, L.V., Lage, M., 2016. Validation of the nursing activities score in portuguese intensive care units. Rev. Bras. Enferm. 69 (5), 826–832. https://doi.org/10.1590/0034-7167-2016-0147.

Miranda, R., Moreno, R., Iapichino, G., 1997. Nine equivalents of nursing manpower use score (NEMS). Intensive Care Med. 23, 760–765.

Miranda, D., Rijk, A., Schaufeli, W., 1996. Simplified therapeutic intervention scoring system: the TISS-28 items - results from a multicenter study. Crit. Care Med. 24 (1), 64–73.

Miranda, D., Nap, R., De Rijk, A., Schaufeli, W., Iapichino, G., Abizanda, R., Agthé, D., Alexander, J.P., Angus, D.C., Artigas, A., Bastos, P., Bihari, D., Boyle, M., Buch, T.F., Burchardi, H., Clarke, T., Dobb, G., Evans, S., Hart, G., Zimmerman, J.E., 2003. Nursing activities score. Crit. Care Med. 31 (2), 374–382. https://doi.org/10.1097/01.CCM.0000045567.78801.CC.

Moghadam, S., Sadoughi, F., Velayati, F., Ehsanzadeh, S.J., Poursharif, S., 2021. The effects of clinical decision support system for prescribing medication on patient outcomes and physician practice performance: a systematic review and meta-analysis. BMC Med. Inf. Decis. Making 21 (1). https://doi.org/10.1186/S12911-020-01376-8.

Mokkink L, Prinsen C, Patrick D, Alonso J, Bouter L, de Vet H, & Terwee C. (2018). *COSMIN manual for systematic reviews of PROMs COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual*. www.cosmin.nl.

Mokkink, L., Prinsen, C., Patrick, D., Alonso, J., Bouter, L., de Vet, H. C., & Terwee, C. (2019). *COSMIN Study Design checklist for Patient-reported outcome measurement instruments*. www.cosmin.nl.

Moreno, R., Morais, P., 1997. Validation of the simplified therapeutic intervention scoring system on an independent database. Intensive Care Medical 23, 640–644.

Padilha, K.G., Stafseth, S., Solms, D., Hoogendoom, M., Monge, F.J.C., Gomaa, O.H., Giakoumidakis, K., Giannakopoulou, M., Gallani, M.C., Cudak, E., Nogueira, L.D.S., Santoro, C., De Sousa, R.C., Barbosa, R.L., Miranda, D.D.R., 2015. Nursing activities score: an updated guideline for its application in the intensive care unit. Rev. Esc. Enferm. U.S.P. 49 (spe), 131–137. https://doi.org/10.1590/S0080-623420150000700019.

Padrón, A., Gutiérrez, C., Enriquez, P., Rivero, Y., 2003. Escala de evaluación de la carga de trabajo en enfermería (EVECTE). Propuesta de una nueva escala. Rev. Cub. Med. Int. y Emerg. 2 (3), 50–58..

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372. https://doi.org/10.1136/bmj.n71.

Palese, A., Comisso, I., Burra, M., Ditaranto, P.P., Peressoni, L., Mattiussi, E., Lucchini, A., 2016. Nursing activity score for estimating nursing care need in intensive care units: findings from a face and content validity study. J. Nurs. Manag. 24 (4), 549–559. https://doi.org/10.1111/JONM.12357.

Porcel-Gálvez, A.M., Allande-Cussó, R., Fernández-García, E., Naharro-Álvarez, A., Barrientos-Trigo, S., 2022. Predictive validity of the INTEGRARE scale in identifying

the risk of hospital-acquired pressure ulcers in acute care hospital settings. J. Tissue Viability 31 (2), 339–345. https://doi.org/10.1016/J.JTV.2022.01.001.

Prescott, P.A., Ryan, J.W., Soeken, K.L., Castorr, A.H., Thompson, K., Phillips, C.Y., 1991. The Patient Intensity for Nursing Index: A Validity Assessment. Researchin Nursing & Health Vol. 14.

Prescott, P., Soeken, K., Ryan, J., 1989. Measuring patient intensity: a reliability study. Eval. Health Prof. 12 (3), 255–269.

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

Queijo, A.F., Martins, R.S., Andolhe, R., Oliveira, E.M., Barbosa, R.L., Padilha, K.G., 2013. Nursing workload in neurological intensive care units: cross-sectional study. Intensive Crit. Care Nurs. 29 (2), 112–116. https://doi.org/10.1016/j.iccn.2012.08.001.

Rayyan. (n.d.). *Rayyan - A Powered Tool for Systematic Literature Reviews*. Retrieved November 28, 2023, from https://www.rayyan.ai/.

Reis Miranda, D., Nap, R., De Rijk, A., Schaufeli, W., Iapichino, G., Abizanda, R., Agthé, D., Alexander, J.P., Angus, D.C., Artigas, A., Bastos, P., Bihari, D., Boyle, M., Buch, T.F., Burchardi, H., Clarke, T., Dobb, G., Evans, S., Hart, G., Zimmerman, J.E., 2003. Nursing activities score. Crit. Care Med. 31 (2), 374–382. https://doi.org/10.1097/01.CCM.0000045567.78801.CC.

Rivera, D.I.C., Sandoval, J.P.R., Zambrano, L.Y.C., Montes, M.A.P., Romero, L.A.L., 2022. Métodos empleados Para cuantificar la carga de trabajo en enfermería en las unidades de cuidados intensivos: una revisión de la literatura. Revista Cuidarte 13 (3), 23011–230113. https://doi.org/10.15649/CUIDARTE.2301.

Rollán, G., Carmona, F., Quirós, C., Cerrillo, I., Jara, A., García, S., Alameda, R., & Martínez, M. (2011). Escalas de medida de carga de trabajo de enfermería en unidades de cuidados críticos. Correlación entre NAS y NEMS. *NURE Investigación*, *8* (55). https://www.nureinvestigacion.es/OJS/index.php/nure/article/view/556.

Ross, P., Howard, B., Ilic, D., Watterson, J., Hodgson, C.L., 2023. Nursing workload and patient-focused outcomes in intensive care: a systematic review. Nurs. Health Sci. https://doi.org/10.1111/NHS.13052.

Sánchez-Sánchez, M.M., Arias-Rivera, S., Fraile-Gamo, M.P., Thuissard-Vasallo, I.J., Frutos-Vivar, F., 2015. Validating the spanish version of the nursing activities score. Enfermeria Intensiva 26 (2), 63–71. https://doi.org/10.1016/j.enfi.2015.02.003.

Schünemann, H., Brozek, J., Guata, G., & Oxman, A. (2013). *GRADE handbook*. Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. https://gdt.gradepro.org/app/handbook/handbook.html.

Silva, A. D., Chavaglia, S. R. R., Ohl, R. I. B., Barbosa, M., Pires, F. C., & Pereira, C. B. de M., 2021. Nurses' workload in critical care units. *Enfermagem Uerj*, *29*, NA-NA. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=01043552&v=2.1&it=r&id=GALE%7CA695144825&sid=googleScholar&linkaccess=fulltext.

Stafseth, S.K., Tønnessen, T.I., Diep, L.M., Fagerstrøm, L., 2018. Testing the reliability and validity of the nursing activities score in critical care nursing. J. Nurs. Meas. 26 (1), 142–162. https://doi.org/10.1891/1061-3749.26.1.142.

Subirana, M., Solà, I., 2006. Revisión sistemática de los instrumentos que miden la actividad de Enfermería y su repercusión sobre los resultados en salud. Metas de Enfermería 9 (6), 22–27.

Toffoletto, M.C., Reynaldos Grandón, K.L., Molina Muñoz, Y., Grillo Padilha, K., 2018. Validation of the nursing activities score in chilean intensive care units. Ciencia y Enfermeria 24. https://doi.org/10.4067/s0717-95532018000100213.

Valls-Matarín, J., Salamero-Amorós, M., 2015. Análisis de la carga de trabajo y uso de los recursos enfermeros en una unidad de cuidados intensivos. Enferm Intensiva 26 (2), 72–81. https://doi.org/10.1016/j.enfi.2015.02.002.

Walther, S.M., Jonasson, U., Karlsson, S., Nordlund, P., Johansson, A., Malstam, J., Malstam, M., South-Eastern, T., 2004. Multicentre study of validity and interrater reliability of the modified nursing care recording system (NCR11) for assessment of workload in the ICU. Acta Anaesthesiol Scandinavica 48, 690–696. https://doi.org/10.1111/j.0001-5172.2004.00397.x.

Wang, Z., Zheng, J., You, L., Wang, Y., Gao, M., Guan, X., 2018. Evaluation of the simplified therapeutic intervention scoring system: chinese version. Intensive Crit. Care Nurs. 45, 85–90. https://doi.org/10.1016/j.iccn.2017.09.009.

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M., 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155 (8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

Wysokiński, M., Ksykiewicz-Dorota, A., Fidecki, W., 2013. Scope of nursing care in polish intensive care units. Biomed Res. Int. 2013 https://doi.org/10.1155/2013/463153.

Yamase, H., 2003. Development of a comprehensive scoring system to measure multifaceted nursing workloads in ICU. Nurs. Health Sci. 5 (4), 299–308. https://doi.org/10.1046/j.1442-2018.2003.00165.x.