



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS ARARANGUÁ
DEPARTAMENTO DE CIÊNCIAS DA SAÚDE
CURSO DE FISIOTERAPIA

LUANA GLENZEL

**Confiabilidade das escalas de dor e comportamento para prematuros: uma revisão
sistemática**

Araranguá
2021

Luana Glenzel

**Confiabilidade das escalas de dor e comportamento para prematuros: uma revisão
sistemática**

Artigo apresentado ao Curso de Graduação em Fisioterapia,
da Universidade Federal de Santa Catarina, como requisito
parcial da disciplina de Trabalho de Conclusão de Curso II.

Orientadora: Profa. Dra. Cristiane Aparecida Moran

Araranguá

2021

DEDICATÓRIA

Dedico esse trabalho à minha família e meus amigos, que me apoiaram e incentivaram durante todo o caminho. Amo vocês!

AGRADECIMENTOS

Primeiramente gostaria de agradecer a Deus, por tudo que Ele fez até aqui, por todas as bênçãos, e por me dar força para continuar mesmo nos momentos difíceis.

Aos meus pais, José e Carla, sem eles eu não estaria aqui, sou muito grata pela vida que me deram. Por terem me proporcionado a oportunidade de estudar sem que eu precisasse me preocupar com outra coisa.

Em especial à minha mãe, Carla, que desde sempre me apoiou e nunca mediu esforços para me ver bem e feliz. Por sempre acreditar em mim e confiar no meu potencial. Por ser meu porto seguro e por todos os momentos de carinho e cuidado. E principalmente por sempre me incentivar e me apoiar em todos os meus sonhos, se hoje estou aqui, com certeza é por ti!

À minha irmã, Laura, por sempre ter me apoiado e acreditado em mim. Obrigada por todo amor e carinho.

Ao meu namorado, Gabriel, por sempre incentivar e por me apoiar em todos os momentos. Obrigada por todo carinho e tempo dedicado a mim, pela tua enorme paciência, companheirismo e amor.

À minha orientadora, Cristiane Aparecida Moran, por todo incentivo e ensinamento e por me fazer ter certeza que a Neonatologia é a área onde está meu coração. Obrigada de verdade!

À minha parceira de revisão sistemática e amiga, Patrícia, muito obrigada de verdade, por todos os momentos de ajuda e ensinamentos. Por todas as conversas e desabafos. Obrigada por estar comigo nesse período, foi muito importante, saiba que tu és uma inspiração pra mim.

Aos meus amigos, Matheus, Juliana e Gustavo, por terem trilhado esse caminho junto comigo. Por me apoiarem e me ajudarem a enfrentar as dificuldades da faculdade. Sem vocês não teria a mesma graça. Obrigada de coração por estarem sempre comigo. Amo vocês.

À minha terapeuta, Cristina que me fez seguir quando eu quis desistir. Você é luz Cris!

E de uma forma geral, a todos que passaram pela minha vida e contribuíram direta ou indiretamente para a minha formação acadêmica.

Confiabilidade das escalas de dor e comportamento para prematuros: uma revisão sistemática

Luana Glenzel^a,
Patricia do Nascimento Oliveira^a,
Cristiane A. Moran^a, PhD

Afiliação: ^aUniversidade Federal de Santa Catarina, Campus Araranguá, Araranguá, Santa Catarina

Endereço de correspondência para: Cristiane A. Moran, Departamento de Ciências da Saúde, Universidade Federal de Santa Catarina.
Rua Governador Jorge Lacerda, Araranguá SC, 3201, [cristiane.moran@ufsc.br], (48) 3721-6255

Título curto: Confiabilidade de escalas para prematuros

Conflitos de interesse: Os autores não têm conflito de interesse a divulgar.

Financiamento/Suporte: Todas as fases deste estudo foram apoiadas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Papel do Financiador/Patrocinador: O CNPq não teve nenhum papel no desenho e condução do estudo.

Abreviações: RN: recém-nascido; UTIN: Unidade de Terapia Intensiva Neonatal; RNPT: recém-nascido pré-termo; COSMIN: padrões baseados em consenso para a seleção de instrumentos de medição de saúde; ABC: ABC pain scale; EVENDOL: Evaluation Enfant Douleur; BIIP: Behavioral Indicators of Infant Pain; BPSN: Bernese Pain Scale for Neonates; NIPS: Neonatal Infant Pain Scale; PASPI: Pain Assessment Scale for Preterm Infants; NNICUPAT: Nepean Neonatal Intensive Care Unit Pain Assessment Tool; NFCS: Neonatal Facial Coding System; NFCS-R: Neonatal Facial Coding System - Revised; N-PASS: Neonatal Pain Agitation and Sedation Scale; PAT: Pain Assessment Tool; ALPS-neo: Astrid Lindgren and Lund Children's Hospitals Pain and Stress; CHIPPS: Children's and Infants Postoperative Pain Scale; PIPP: Premature Infant Pain Profile; PIPP-R: Premature Infant Pain Profile - Revised; CSS: Clinical Scoring System; GIVES: Neonatal Pain Assessment Scale; FANS: Faceless Acute Neonatal Pain Scale; EDIN: Échelle Douleur Inconfort Nouveau-Né; COVERS: neonatal pain scale; NIAPAS: Neonatal Infant Acute Pain Assessment Scale; CPSPN: Clinical Pain Scale for Preterm Neonates; ICC: Coeficiente de Correlação Intraclasse.

Resumo do Artigo

Revisão sistemática sobre a confiabilidade das escalas que avaliam dor, estresse e/ou comportamento dos recém-nascidos prematuros internados na unidade de terapia intensiva neonatal.

Depoimento dos contribuintes

Dra. Cristiane Aparecida Moran, Luana Glenzel e Patricia do Nascimento Oliveira conceituaram e projetaram o estudo.

Dra Cristiane Aparecida Moran coordenou, supervisionou e revisou o estudo.

Luana Glenzel e Patricia do Nascimento Oliveira redigiram, revisaram e realizaram todas as etapas do estudo.

Resumo

Contexto: Existem diversas escalas que avaliam comportamento, dor e/ou estresse em recém nascidos pré-termos internados na UTIN, entretanto há uma lacuna científica sobre qual a ferramenta ideal.

Objetivo: Revisar sistematicamente ferramentas utilizadas na avaliação do comportamento, estresse e/ou dor de recém-nascidos prematuros nas UTIN e demonstrar a confiabilidade dessas ferramentas a partir da ferramenta COSMIN.

Fontes de dados: Pesquisas eletrônicas foram conduzidas no PubMed/MEDLINE, Cochrane Library, Web of Science, Scientific, SciELO, LILACS e EBSCOhost Research Platform. No total 8.570 registros foram encontrados

Seleção dos estudos: Estudos foram incluídos se avaliassem a confiabilidade das escalas de dor, estresse e/ou comportamento de prematuros internados na UTIN. No total, 46 estudos foram revisados para inclusão.

Extração de dados: Os dados de 32 estudos preencheram os critérios de inclusão e exclusão. Os estudos foram sintetizados pelos resultados da COSMIN.

Resultados: Foram identificadas 24 escalas que avaliaram comportamento, dor, e/ou estresse em RNPT. A confiabilidade inter-examinador das escalas variou de 0.10 a 1.00. A confiabilidade intra-examinador variou de 0.97 a 1.00. A maioria dos estudos demonstrou confiabilidade inter-examinador maior que 0.90 e somente 2 estudos avaliaram confiabilidade inter e intra-examinador.

Limitações: A falta de padronização das escalas e as faixas etárias variáveis na classificação da prematuridade desafiam a prática clínica baseada em provas científicas.

Conclusão: Foram destacadas as escalas BIPP, PIPP-R, EVENDOL e PASPI devido a avaliação da confiabilidade e aplicação da COSMIN.

Abstract

Context: There are several scales that assess behavior, pain and/or stress in preterm newborns admitted to the NICU, however, there is a scientific gap about which tool is ideal.

Objective: To systematically review tools used to assess behavior, stress and/or pain in preterm newborns in the NICU and to demonstrate the reliability of these tools based on COSMIN.

Data Sources: Electronic searches were conducted in PubMed/MEDLINE, Cochrane Library, Web of Science, Scientific, SciELO, LILACS and EBSCOhost Research Platform. In total 8,570 records were found.

Study Selection: Studies were included if they evaluated the reliability of pain, stress and/or behavior scales in premature infants admitted to the NICU. A total of 46 studies were reviewed for inclusion.

Data extraction: Data from 32 studies met the inclusion and exclusion criteria. Studies were synthesized by COSMIN scores.

Results: We identified 24 scales that assessed behavior, pain, and/or stress in PIs. Inter-examiner reliability of the scales ranged from 0.10 to 1.00. Intra-examiner reliability ranged from 0.97 to 1.00. Most studies demonstrated inter-examiner reliability greater than 0.90 and only 2 studies evaluated both inter and intra-examiner reliability.

Limitations: The lack of standardization of scales and varying age ranges in the classification of prematurity challenges clinical practice based on scientific evidence.

Conclusion: The BIPP, PIPP-R, EVENDOL and PASPI scales were highlighted due to reliability evaluation and COSMIN application.

INTRODUÇÃO

Antigamente a dor nos recém-nascidos pré-termos (RNPT) era subestimada,¹ causando quadros nocivos durante os procedimentos cirúrgicos e invasivos, apesar dos avanços sobre a fisiologia das respostas à dor, poucos pesquisadores questionam os efeitos de métodos e técnicas realizadas nas unidades de terapia intensiva neonatal (UTIN).²

Hatfield et al³ descreveram que a transmissão e percepção de dor estão presentes e desenvolvidas a partir da 23^a semana de gestação, sendo que os neurotransmissores excitatórios estão aumentados e os inibitórios diminuídos, potencializando assim as respostas aos estímulos dolorosos nos prematuros quando comparados com bebês mais velhos e pessoas adultas.

Existem implicações da potencialização da dor no desenvolvimento neonatal,⁴ uma vez que nos dias atuais a inovação tecnológica tem possibilitado uma maior sobrevivência de recém-nascidos prematuros extremos internados UTIN,⁵ no entanto, ainda permeia as condições insalubres sobre a exposição em um ambiente adaptado para representar a vida intrauterina.⁶

Durante a permanência hospitalar os RNPT são submetidos a inúmeros procedimentos diagnósticos e terapêuticos estressantes ou dolorosos.⁵ Em virtude da imaturidade do sistema nervoso, a exposição repetida a esses eventos pode acarretar efeitos adversos, um exemplo é a hipersensibilidade à dor, causando sofrimento nos procedimentos supracitados. Ademais, a exposição aos eventos dolorosos pode influenciar no desenvolvimento infantil e as mudanças na sensibilidade à dor podem permanecer após o período neonatal. As consequências em longo prazo podem variar desde a mudanças na sensibilidade à dor até a mudanças neuroanatômicas, comportamentais, emocionais e dificuldade de aprendizagem.^{2,7}

Quando os recém-nascidos (RN) são expostos a situações ou procedimentos que causam dor e/ou estresse, o organismo dos prematuros libera hormônios como adrenalina, noradrenalina

e cortisol. Substâncias que irão produzir altos níveis de estresse e modificações em diferentes sistemas, causando respostas como apneias, diferenças na frequência cardíaca, pressão arterial e diminuição da saturação de oxigênio.⁸

A avaliação da resposta comportamental do RNPT, para que seja possível identificar os procedimentos que lhe causam dor e/ou estresse durante a internação hospitalar, se torna fundamental para diminuir o desconforto causado por procedimentos realizados nas UTIN. Atualmente, existem algumas escalas que auxiliam nessa avaliação. No entanto, há uma lacuna científica sobre qual a ferramenta ideal para avaliar eventos dolorosos e estressantes durante a internação hospitalar.⁹

Assim, o presente estudo tem como objetivo realizar uma busca sistemática sobre as ferramentas utilizadas para avaliar o comportamento, estresse e/ou dor de RNPT nas UTIN e demonstrar a confiabilidade dessas ferramentas a partir da ferramenta padrões baseados em consenso para a seleção de instrumentos de medição de saúde (COSMIN).

MÉTODOS

O estudo foi desenhado como uma revisão sistemática da literatura, sendo realizada uma estratégia de busca sobre as escalas disponíveis utilizadas para avaliação comportamental/estresse e dor em RNPT, com o objetivo de determinar a confiabilidade dessas ferramentas. Essa revisão sistemática seguiu as recomendações do Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).¹⁰ O protocolo deste estudo foi previamente cadastrado no Registro Prospero Internacional de Revisões Sistemáticas (PROSPERO) em abril de 2021 sob o número de identificação: CRD42021242128.

Estratégia de busca

A coleta de dados foi realizada no período de janeiro de 2021 a março de 2021, nas bases de dados: PubMed/MEDLINE, Cochrane Library, Web of Science, Scientific Electronic Library Online - SciELO, Literatura Latino-Americana e do Caribe em Ciências da Saúde - LILACS e EBSCOhost Research Platform. A seleção dos descritores foi baseada na estratégia PICO descrita na Tabela 1 e nas propriedades de medição identificadas pela COSMIN (Mokkink *et al.*, 2020). Os descritores de saúde utilizados foram: (*Infant, Premature AND Intensive Care Units, Neonatal AND Behavior Rating Scale AND Pain Measurement AND Infant Behavior AND Stress, Physiological AND Validation Study AND Cross-Cultural Comparison AND Reproducibility of Results AND Bias*).

CrITÉRIOS de inclusão e exclusão

Os critérios de inclusão foram estudos originais nos idiomas Inglês, Espanhol ou Português que utilizaram ferramentas para avaliar o comportamento, estresse e/ou dor em recém-nascidos pré-termo na UTIN, e que avaliaram a confiabilidade das ferramentas de medição. E os critérios de exclusão foram estudos não disponíveis em texto completo, artigos publicados antes do ano 2000, revisões sistemáticas, monografias, teses, dissertações e resumos de congressos científicos.

Seleção dos estudos

A seleção dos artigos foi realizada no aplicativo Mendeley Desktop e conduzida por dois revisores (LG e PNO) de forma independente, de acordo com os critérios de inclusão e nos casos de discordância foi consultada a pesquisadora (CAM). Inicialmente foram excluídos os estudos

duplicados e em seguida os artigos foram selecionados pela leitura de título e resumo. Após isso, os artigos selecionados foram lidos na íntegra para avaliar a elegibilidade. Por último, foi realizada uma busca manual a partir das referências encontradas, para garantir a inclusão de todos os artigos disponíveis.

Extração dos dados e risco de viés

Dois revisores (LG e PNO) extraíram os dados individualmente e às cegas. Os dados selecionados dos estudos foram sobre a caracterização da amostra (número, idade gestacional, sexo e idade pós-natal), sobre a escala de avaliação utilizada e o construto a ser medido, sobre dados básicos dos estudos como (autor, ano e desenho de estudo) e por fim sobre os dados utilizados pelos estudos para avaliação da confiabilidade das ferramentas de medição. Os dados foram inseridos em uma planilha no programa Microsoft Excel[®].

A avaliação do risco de viés foi realizada e todos os estudos incluídos foram revisados utilizando os padrões baseados em consenso para a seleção de instrumentos de medição de saúde (COSMIN). A ferramenta COSMIN foi desenvolvida para avaliar a qualidade metodológica de estudos sobre propriedade de medição.¹¹ A ferramenta avalia diferentes propriedades de medição a partir de padrões específicos separados em 10 caixas. Para avaliar a qualidade dos artigos incluídos nesta revisão utilizamos a caixa 1 (desenvolvimento PROM), caixa 4 (consistência interna), caixa 6 (confiabilidade) e caixa 9 (teste de hipóteses para validade do construto). Cada uma possui um sistema de classificação de quatro pontos, sendo “muito bom”, “adequado”, “duvidoso” e “inadequado” ou “não aplicável”. Os itens da ferramenta COSMIN utilizados neste estudo estão descritos na Tabela 2.

Análise de dados

Os valores do coeficiente Kappa são interpretados como 0.00 a 0.20 indicando concordância leve, 0.21 a 0.40 concordância razoável, 0.41 a 0.60 concordância moderada, 0,61 a 0,80 concordância substancial e de 0.81 a 1.00 indica concordância quase perfeita ou perfeita.^{12,13} O Coeficiente de Correlação Intraclasse - ICC pode ser classificado como baixa confiabilidade para valores abaixo de 0.40, média confiabilidade de 0.40 a 0.59, boa confiabilidade para valores entre 0.60 a 0.74 e excelente confiabilidade para valores de 0.75 a 1.0.¹³

O coeficiente de Pearson é classificado como muito forte quando relata valores acima de 0.90, valores entre 0.70 e 0.90 são classificados como altos e valores entre 0.50 e 0.70 são classificados como moderados.^{14,15} O alfa de Cronbach para ser considerado como aceitável deve ser acima de 0.70.¹⁶ O valor máximo do alfa de Cronbach é 0.90.¹⁷

RESULTADOS

Busca

Inicialmente, foram encontrados 8570 artigos, sendo 35 incluídos pela busca manual, após a aplicação dos critérios de inclusão e exclusão foram selecionados 46 artigos para serem lidos na íntegra, destes 13 foram excluídos por incompatibilidade dos critérios de inclusão (3) RNT, (8) não avaliou confiabilidade, (1) protocolo de estudo, (1) UTI pediátrica e (1) indisponível em texto completo. Assim resultando em trinta e dois artigos selecionados. A estratégia de busca e do processo de seleção dos estudos estão descritos no fluxograma Figura 1.

Características dos estudos

Foram incluídos 2623 indivíduos sendo 51,8% do sexo masculino, a idade gestacional variou de 23 a 42 semanas, a idade pós natal variou de 0 dias a 18 meses. 10 artigos avaliaram as escalas apenas em RNPT,^{25,27,30-32,35,46,47,50,52} e um artigo avaliou também em lactentes.³⁴ Os outros 21 artigos aplicaram as escalas em RNPT e RNT.^{26,28,29,33,36-45,48,49,51,53,54,55,56}

Dos 32 artigos selecionados, foram identificadas 24 escalas para avaliação da dor, estresse e/ou comportamento em RN pré-termo, dessas, onze são unidimensionais e treze são multidimensionais. As escalas unidimensionais apresentam somente um indicador de avaliação de dor, podendo ser fisiológicos ou comportamentais e as escalas multidimensionais são mais abrangentes, pois apresentam aspectos fisiológicos e comportamentais,¹⁸ as informações das escalas estão demonstradas na Tabela 3.

Confiabilidade

Os dados referentes à confiabilidade das escalas estão apresentados na Tabela 4. Dos 32 estudos, somente um não avaliou a confiabilidade inter-examinador.⁵¹ Dois artigos avaliaram tanto a confiabilidade inter-examinador quanto a confiabilidade intra-examinador.^{42,43} Oito estudos não realizaram a avaliação da consistência interna.^{29-31,34,37,43,50,56}

A confiabilidade inter-examinador das escalas variou de 0.10²⁹ a 1.00.^{34,39,52} A confiabilidade intra-examinador variou de 0.97⁴³ a 1.00⁴². Vinte e três estudos demonstraram confiabilidade inter-examinador maior que 0.90,^{26-28,32-39,41-44,46-49,55-55} a consistência interna variou de 0.67⁵⁴ a 0.99.³⁶ Dois estudos avaliaram teste-reteste^{36,55} e outros dois avaliaram a sensibilidade das escalas.^{39,45}

Para confiabilidade inter-examinador 16 escalas ficaram acima dos valores relatados como excelente para ICC: EDIN, ALPS-neo, NFCS-R, CHIPPS, PIPP-R, NIPS, N-PASS, BPSN, FANS, BIIP, NIAPAS, COVERS, PAT, COMFORT adaptada, PASPI e CPSPN.^{25,26,28,32,35,36,39-41,44,46-49,52,54-56} As escalas EVENDOL e NFCS também tiveram

confiabilidade inter-examinador quase perfeita, porém foram avaliadas com o coeficiente Kappa ponderado.^{27,34} Um estudo avaliou confiabilidade inter-examinador e intra-examinador da escala BPSN pelo alfa de Cronbach, obtendo valores acima de 0.70.⁴³

Um estudo avaliou a confiabilidade inter-examinador da escala NIAPAS a partir do coeficiente Pearson, obtendo um valor muito forte igual a 0.99.⁴² O coeficiente Pearson também foi utilizado para avaliar a confiabilidade inter-examinador das escalas NIPS, PIPP e PIPP-R, nesses estudos as escalas obtiveram valores altos, ou seja, acima de 0.70.^{37,38,53}

Para consistência interna 18 escalas obtiveram valores de alfa de Cronbach entre 0.70 a 0.90, o que é considerado adequado, as escalas foram: EDIN, EVENDOL, CHIPPS, PIPP-R, NIPS, N-PASS, BPSN, FANS, BIIP, PIPP, NIAPAS, COVERS, PAT, COMFORT-neo, COMFORT adaptada, PASPI, GIVENS e CPSPN.^{27,28,32,33,35,36,38-42,44-49,51,53,54}

Cinco escalas EDIN, ALPS-neo, NFCS-R, NIAPAS e CPSPN tiveram valores de alfa de Cronbach maior que 0.90^{25,26,38,53,55} e uma escala, PIPP-R, para medida em estresse obteve valores abaixo de 0.70 para consistência interna.⁵⁴ Apenas para duas escalas BIIP e NIAPAS foi realizado teste reteste e ambas obtiveram valores acima de 0.90,^{36,55} a sensibilidade foi avaliada em dois estudos para as escalas PIPP-R e COMFORT-neo e obtiveram os valores 91.5% e 81% respectivamente.^{39,45}

Qualidade dos Estudos

Os dados de qualidade de cada estudo estão demonstrados Figura 2. Acerca dos padrões e itens utilizados, na caixa 1, no item 1, 71,9%^{25-28,31-38,41,44,46-50,52-55} dos artigos foram classificados como “muito bons” e 28,1%^{29,30,39,40,42,43,45,51,56} foram classificados como “inadequados”, demonstrando que a maioria dos artigos traz em seus estudos uma descrição clara do construto a

ser medido. No item 4, 34,4%^{26,27,31,36,41-45,53} dos artigos foram classificados como “muito bons”, enquanto 65,6%^{25,28-30,33-35,37-40,46-52,54-56} foram classificados como “duvidosos”, o que significa que a maioria dos estudos não forneceu uma descrição clara do contexto de uso das escalas. Para o item 13, 28,1%^{36-39,44,45,51,56} dos artigos foram classificados como “muito bons”, 37,5%^{25-27,30,32,33,35,40,46,47,49,54} foram classificados como “adequados”, 21,9%^{28,29,31,34,41,42,48} foram classificados como “duvidosos” e 12,5%^{43,50,52,53} foram classificados como “inadequados”, indicando que mais de 50% dos estudos tiveram um bom tamanho de amostra. Em relação a caixa 4, no item 2, 78,1%^{25-28,32,33,35,36,38-55} dos artigos foram classificados como “muito bons”, enquanto 12,5%^{31,37,56} foram classificados como “inadequados” e o restante como não aplicável,^{29,30,34} mostrando que a consistência interna foi avaliada na maioria dos artigos.

Sobre a caixa 6, no item 1, poucos estudos tiveram uma boa classificação, 15,6%^{27,44,47,48,50} dos estudos foram classificados como “muito bons” e 18,8%^{29,32,35,38,42,46} foram classificados como “adequados”, o restante (65,6%)^{25,26,28,30,31,33,34,36,37,39-41,43,45,49,51-56} foram classificados como “duvidosos”, indicando que mais da metade dos estudos não trouxeram informações acerca da estabilidade clínica dos pacientes. Para o item 4, a maioria dos estudos administraram a medição sem conhecimento dos escores ou valores de outras medições repetidas nos mesmos pacientes, pois, 31,3%^{27,32,34-36,45,47,50,53,54} dos estudos foram classificados como “muito bons”, 40%^{26,28,29,30,37-44,52} foram classificados como “adequados”, 25% classificados como “duvidosos”^{25,33,46,48,49,51,55,56} e somente 1 artigo foi classificado como “inadequado”.³¹ Para o item 5, 46,9%^{29,31-34,36,39,42-44,47,50,53-55} dos estudos foram classificados como “muito bons”, 18,8%^{25,27,28,30,35,37} foram classificados como “adequados”, ou seja, a maioria dos estudos demonstrou que os profissionais atribuíram os escores ou determinaram valores sem conhecimentos de medições repetidas nos mesmos pacientes, porém alguns artigos

(34,4%)^{26,38,40,41,45,46,48,49,51,52,56} foram classificados como “duvidoso”, pois essa informação estava obscura.

Ainda sobre a caixa 6, no item 6, somente 6,3%^{36,37} dos artigos foram classificados como “muito bons”, o restante (84%)^{25-28,31-35,38-50,52-56} foi classificado como “duvidoso” e 9,4% foram “inadequado”,^{29,30,51} evidenciando que a maioria dos estudos não esclareceu se os avaliadores eram experientes, se foram previamente treinados e não relatou os critérios de inclusão e exclusão utilizados. Para o item 7, 12,5%^{35,36,41,48} dos estudos foram classificados como “muito bons”, 50%^{26-28,32,33,39,40,44,46,47,49,50,52,54-56} foram classificados como “adequados”, 15,6%^{31,37,38,42,53} foram classificados como “duvidosos”, 6,3%^{43,51} foram classificados como “inadequado” e para 15,6%^{25,29,32,34,45} como “não aplicável”, demonstrando que a maioria dos estudos realizou a avaliação do ICC, porém não apresentou a fórmula. No item 8, 9,4%^{34,45,46} classificados como “muito bons”, 15,6%^{25-27,33,55} foram classificados como “adequados” e 75%^{28-32,35-44,47-54,56} dos estudos foram classificados como “não aplicável”, pois o kappa ponderado não foi utilizado. Para o item 9, 9,4%^{29,30,40} dos estudos foram classificados como “muito bons” e o restante (90,6%)^{25-28,31-39,41,56} foram classificados como “não aplicável”, pois não apresentaram valores de kappa. Em relação a caixa 9, item 1, 53,1%^{27,28,32-36,41,42,46-48,50,52,55} dos artigos foram classificados como “muito bons”, 9,4%^{43,51,56} foram classificados como “inadequados” e 37,5%^{25,26,29,37-40,44,45,49,53,54} como “não aplicável”, demonstrando que a maioria dos estudos que utilizou de um instrumento comparador, descreve de forma clara o que o era medido pelo instrumento.

Escalas

Das 24 escalas encontradas em nosso estudo, as escalas EVENDOL, CHIPPS, PIPP-R,

N-PASS, BPSN, FANS, BIIP e PASPI obtiveram avaliação classificada como ótima, tanto para confiabilidade inter-avaliador quanto para consistência interna. Em relação aos artigos incluídos nesse estudo, os estudos de Chen et al³⁶, Sadeghi et al⁴⁴, Tristão et al²⁷ e Liaw et al⁴⁷ demonstraram melhor qualidade e menor risco de viés.

Somando as características de melhor confiabilidade das escalas e melhor qualidade dos estudos, temos que as escalas BIPP, PIPP-R, EVENDOL e PASPI foram consideradas com maior confiabilidade e menor risco de viés, sendo então, destacadas para aplicabilidade clínica. O construto a ser medido, as variáveis e a pontuação total de cada escala estão descritas na tabela 5.

Premature Infant Pain Profile-Revised (PIPP-R)

Os indicadores fisiológicos e faciais formam o subtotal, se nesses itens a pontuação for maior que zero, deve-se adicionar as pontuações dos indicadores de idade gestacional (IG) e estado comportamental (EC). Para pontuação total, será somado a pontuação do subtotal (indicadores fisiológicos e faciais), mais a pontuação da IG e mais a pontuação de EC.⁵⁷

A escala PIPP-R deve ser pontuada da seguinte maneira:

Passo 1: observar o RN por 15 segundos, em repouso e avaliar os sinais vitais [frequência cardíaca (FC) mais alta, saturação de oxigênio (SatO₂) mais baixa e EC]. Passo 2: observar o RN por 30 segundos após o procedimento e avaliar a mudança dos indicadores (FC mais alta, SatO₂ mais baixa e duração das ações faciais), se o RN precisar de aumento da oferta de O₂ em qualquer momento, antes ou durante o procedimento, ele recebe +3 pontos no indicador de SatO₂. Passo 3: pontuar IG e EC se o subtotal for maior que zero. Passo 4: calcular a pontuação total adicionando o subtotal, mais IG, mais EC.⁵⁷

Evaluation Enfant Douleur - EVENDOL

Para realização da pontuação da escala EVENDOL, o RN deve ser observado por 30 segundos imediatamente após o procedimento. No repouso, observar o RN a distância antes de realizar o exame físico ou procedimento. Durante a mobilização deve-se avaliar a dor durante o exame, movimentação, ou palpação da área dolorosa. Reavaliar a dor regularmente após a administração de analgésicos, esperar entre 30 a 45 minutos se o analgésico for administrado via oral ou retal, de 5 a 10 minutos se for administrado por via intravenosa. Anotar se o RN está em repouso ou movimentação.²⁸

Pain Assessment Scale for Preterm Infants - PASPI

A linha de base dos parâmetros fisiológicos para cada RN individual deve ser calculada antes da avaliação da dor e registrada para estabelecer o padrão de cada RN.⁴⁸

DISCUSSÃO

A presente revisão sistemática reuniu 32 estudos que relataram dados sobre a confiabilidade das escalas que avaliam dor, comportamento e/ou estresse em recém-nascidos pré-termos internados na UTIN. Como resultado foram encontradas 24 escalas validadas para o uso na população em questão. As escalas EVENDOL, CHIPPS, PIPP-R, N-PASS, BPSN, FANS, BIIP e PASPI obtiveram avaliação classificada como ótima, tanto para confiabilidade inter-avaliador quanto para consistência interna e perante nossos resultados foram consideradas relevantes clinicamente.

Para a confiabilidade inter-examinador avaliada pelo ICC as escalas CHIPPS, PIPP-R, N-PASS, BPSN, FANS, BIIP e PASPI obtiveram uma pontuação considerada excelente. Na

confiabilidade inter-avaliador avaliada pelo Kappa ponderado a escala EVENDOL pontuou um valor considerado muito bom. Existem alguns fatores que podem influenciar no melhor ou pior valor da confiabilidade inter-avaliador, como a qualidade dos estudos, que pode ser influenciada pela presença ou não de experiência dos avaliadores e pela realização ou não de treinamento prévio.¹⁹ Outro fator que pode influenciar são as variáveis contidas nas escalas, pois as variáveis fisiológicas tendem a ter uma melhor correlação com a dor quando comparadas com as variáveis comportamentais, que podem ser diferentes devido a idade gestacional ou gravidade da doença.²⁰

Além disso, nossos resultados demonstraram variados modos de avaliação da confiabilidade inter-examinador: coeficiente Kappa, Kappa ponderado, coeficiente de correlação intraclassa (ICC), coeficiente Pearson e alfa de Cronbach. O coeficiente Kappa é normalmente aplicado quando as variáveis são categóricas nominais, o Kappa ponderado é usado para dados categóricos ordinais e o ICC é normalmente utilizado para dados contínuos. O coeficiente Kappa e o ICC são os mais adequados e utilizados para avaliação da confiabilidade inter-examinador.^{13,21,22}

Com relação à consistência interna, todas as escalas encontradas em nossos resultados foram avaliadas com o alfa de Cronbach e a melhor ou pior pontuação encontrada em nossos resultados pode estar relacionada ao fato de algumas escalas serem unidimensionais ou multidimensionais. Quando as escalas apresentam um modelo multidimensional a alta consistência interna se torna mais difícil de ser alcançada.²⁵ Por mais que as escalas multidimensionais possam parecer melhores para avaliação da dor devido a complexidade que existe em relação à resposta fisiológica, a discordância entre os indicadores comportamentais e fisiológicos podem dificultar o reconhecimento do quadro algico. Os parâmetros fisiológicos, mesmo que valiosos para fornecer medidas objetivas da dor, acabam tendo baixa especificidade,

pois podem ser alterados por outros estímulos não dolorosos e os parâmetros comportamentais podem ser muito subjetivos, o que também dificulta a indicação da dor.^{20,23}

Em nossos resultados, alguns estudos demonstraram alfa de Cronbach maior que 0.90 para as escalas EDIN, ALPS-neo, NFCS-R, NIAPAS e CPSPN.^{25,26,28,52,55} Contudo, esse valor não é esperado, uma vez que escalas que possuem uma consistência interna maior do que 0.90 mostram que suas variáveis podem ser redundantes, no caso, diferentes variáveis estão avaliando os mesmos elementos do construto.¹⁷

Segundo Souza et al²¹ a confiabilidade refere-se à estabilidade, consistência interna e equivalência de uma medida. A estabilidade mede a semelhança dos resultados obtidos em dois momentos diferentes e pode ser realizada pelo teste reteste ou em casos de variáveis contínuas pelo ICC. A consistência interna demonstra se todas as variáveis de um instrumento medem a mesma característica, a medida mais utilizada para avaliação da consistência interna é o coeficiente alfa de Cronbach. A equivalência refere-se ao grau de concordância entre dois ou mais observadores ao que diz respeito às pontuações de um instrumento e pode ser medida através do coeficiente Kappa ou ICC. Em nosso estudo nenhuma das escalas encontradas foram avaliadas em todos esses quatro domínios, mas sabe-se que a escolha dos testes estatísticos que são usados para avaliar a confiabilidade pode variar dependendo do que se pretende medir.²¹

No que diz respeito às escalas que obtiveram uma excelente confiabilidade inter-examinador e boa consistência interna neste estudo, nossos resultados corroboram com Giordano et al²⁴ para algumas escalas encontradas, pois eles também demonstram em seu estudo que as escalas CHIPPS, BIIP, EVENDOL, N-PASS, foram consideradas adequadas perante suas avaliações. Ademais, perante os resultados de Giordano et al²⁴ as escalas PIPP-R, BPSN, FANS e PASPI não foram consideradas relevantes, o que difere dos nossos achados, tendo em vista que

essas escalas demonstraram-se adequadas devido a excelente confiabilidade inter-examinador e boa consistência interna.

A escala PIPP-R a partir de nossos resultados foi considerada adequada, porém não obteve uma boa consistência interna quando o construto a ser medido é estresse, fato este que pode ser explicado devido a PIPP-R ter sido desenvolvida primeiramente para avaliar a dor.⁵⁴

As escalas ABC, PAT, NIAPAS, COMFORT-adaptada, COVERS, NFCS, NIPS, PIPP, COMFORT-neo e EDIN foram consideradas adequadas clinicamente perante os resultados de uma revisão sistemática conduzida por Giordano et al²⁴. Em nossos resultados, essas escalas não foram consideradas adequadas pois não obtiveram valores considerados ótimos ou excelentes em relação ao ICC ou Kappa, não realizaram avaliação da consistência interna, foram consideradas redundantes devido a valores acima de 0,90 de consistência interna e obtiveram variada diferença em relação a diferentes estudos que avaliaram a mesma escala.

Em relação a escala BIIP, Giordano et al²⁴ demonstraram em seus resultados que a escala foi considerada com certo nível de risco de viés devido a amostra não ser do tamanho considerado adequado, o que difere de nossos resultados, tendo em vista que um dos estudos incluídos nessa revisão que avaliou a escala BIPP obteve o melhor desempenho em relação ao risco de viés avaliado pela COSMIN, entre esses, o tamanho da amostra que foi classificado como muito bom.³⁶ O que indica que os artigos mais recentes conseguiram superar a limitação existente anteriormente.

Ao que diz respeito a qualidade dos estudos, nossos resultados demonstraram que através da avaliação pela COSMIN, além de Chen et al³⁶ outros três estudos ultrapassaram 50% os itens classificados como “muito bons” e não apresentaram itens classificados como “inadequados”.^{27,44,47} Assim, o estudo de Sadeghi et al⁴⁴ que avaliou a escala PIPP-R, apresentou

como duvidoso apenas o item “houve alguma falha importante no projeto ou nos métodos estatísticos do estudo?” pois ele não mencionou sobre treinamento e experiência prévia dos avaliadores. Já Tristão et al²⁷ avaliaram a confiabilidade da escala EVENDOL e obteve como resultado duvidoso o item “houve alguma falha importante no projeto ou nos métodos estatísticos do estudo?” porque não relata o tempo de experiência dos avaliadores. E o estudo de Liaw et al⁴⁷ avaliou a escala PASPI e foi classificado como duvidoso pelos itens “houve alguma falha importante no projeto ou nos métodos estatísticos do estudo?” e “é fornecida uma descrição clara do contexto de uso?” pois não apresentou o tempo de experiência dos avaliadores e na descrição do artigo não ficou claro qual é o contexto em que a escala pode ser utilizada.

Por fim, com base na presente revisão sistemática, perante as avaliações de confiabilidade inter-examinador, consistência interna e qualidade dos estudos pela COSMIN, as escalas BIPP, PIPP-R, EVENDOL e PASPI foram as escalas consideradas adequadas e com menor risco de viés, sua aplicação é voltada para RNPT internados na UTIN e o construto a ser avaliado são comportamento (BIPP) e dor aguda (PIPP-R, EVENDOL e PASPI).

LIMITAÇÕES

Esta revisão sistemática fornece uma visão sobre as escalas usadas na UTIN em RNPT, no entanto, a falta de padronização das escalas e as variáveis faixas etárias na classificação da prematuridade ainda desafiam a prática clínica baseada em evidência científica.

CONCLUSÃO

De acordo com os resultados encontrados no presente estudo, foram observadas uma variedade de escalas para população de neonatos pré-termos utilizadas no ambiente da UTIN e a

confiabilidade dessas escalas foi avaliada de inúmeras formas, demonstrando diferentes níveis de confiabilidade e os estudos incluídos demonstraram diferentes níveis em relação a qualidade e risco de viés. Por fim, destacamos as escalas BIPP, PIPP-R, EVENDOL e PASPI devido a melhor confiabilidade inter-examinador e consistência interna, além da melhor qualidade e menor risco de viés dos estudos que realizaram a avaliação dessas escalas.

REFERÊNCIAS

1. Anand, KJS, Hickey PR. Pain and its effects in the human neonate and fetus. *N Engl J Med* 1987;317(21):1321-1329.
2. Bucea, O, Riddell RP. Non-pharmacological pain management in the neonatal intensive care unit: managing neonatal pain without drugs. *Seminars in Fetal and Neonatal Medicine* 2019;24(4):101017. doi: 10.1016/j.siny.2019.05.009.
3. Hatfield LA, Murphy N, Karp K, Polomano RC. A Systematic Review of Behavioral and Environmental Interventions for Procedural Pain Management in Preterm Infants. *J Pediatr Nurs*. 2019;44:22-30. doi: 10.1016/j.pedn.2018.10.004.
4. Perry M, Tan Z, Chen J, et al. Neonatal Pain: Perceptions and Current Practice. *Crit Care Nurs Clin North Am*. 2018;30(4):549-561. doi: 10.1016/j.cnc.2018.07.013.
5. Cong X, Wu J, Vittner D, et al. The impact of cumulative pain/stress on neurobehavioral development of preterm infants in the NICU. *Early Hum Dev*. 2017;108:9-16. doi: 10.1016/j.earlhumdev.2017.03.003.
6. Simons SH, van Dijk M, Anand KS, et al. Do we still hurt newborn babies? A prospective study of procedural pain and analgesia in neonates. *Arch Pediatr Adolesc Med*. 2003;157(11):1058-64. doi: 10.1001/archpedi.157.11.1058.
7. Olsson E, Ahl H, Bengtsson K, et al. The use and reporting of neonatal pain scales: a systematic review of randomized trials. *Pain*. 2021;162(2):353-360. doi: 10.1097/j.pain.0000000000002046.
8. Torres-Ake EA, del Rosario Lugo-Ake G, Matos-Villanueva JA et al. Masaje frente a musicoterapia para reducir el estrés en prematuros de una unidad crítica neonatal, una revisión sistemática. *Revista de Enfermería del Instituto Mexicano del Seguro Social*. 2020;28(1):49-57.
9. Maxwell LG, Fraga MV, Malavolta CP. Assessment of Pain in the Newborn: An Update. *Clin Perinatol*. 2019;46(4):693-707. doi: 10.1016/j.clp.2019.08.005.
10. Moher D, Shamseer L, Clarke M, et al. PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1. doi: 10.1186/2046-4053-4-1.
11. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-1179. doi: 10.1007/s11136-017-1765-4.
12. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74. PMID: 843571.
13. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi: 10.20982/tqmp.08.1.p023.
14. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg*. 2018;126(5):1763-1768. doi: 10.1213/ANE.0000000000002864.
15. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in

- medical research. *Malawi Med J.* 2012;24(3):69-71. PMID: 23638278.
16. Nunnally JC. *Psychometric Theory.* 2.ed. New York; McGraw-Hill; 1978.
 17. Almeida D, dos Santos MAR, Costa AFB. Aplicação do coeficiente alfa de Cronbach nos resultados de um questionário para avaliação de desempenho da saúde pública. XXX Encontro Nacional de Engenharia de Produção. 2010;1-12.
 18. Melo, GM, Lélis, ALPA, de Moura AF, et al. Pain assessment scales in newborns: integrative review. *Revista Paulista de Pediatria.* 2014;32(4):395-402. doi: 10.1590/S0103-05822014000400017.
 19. Oliveira ACASF, Scarmagnani RH, Fukushiro AP, et al. Influência do treinamento dos avaliadores no julgamento perceptivo da hipernasalidade. *CoDAS.* 2016;141-148. doi: 10.1590/2317-1782/20162015163.
 20. Hatfield LA, Ely EA. Measurement of acute pain in infants: a review of behavioral and physiological variables. *Biol Res Nurs.* 2015;17(1):100-11. doi: 10.1177/1099800414531448.
 21. Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol Serv Saude.* 2017;26(3):649-659. doi: 10.5123/S1679-49742017000300022.
 22. Mandrekar JN. Measures of interrater agreement. *J Thorac Oncol.* 2011;6(1):6-7. doi: 10.1097/JTO.0b013e318200f983.
 23. Arias MC, Guinsburg R. Differences between uni-and multidimensional scales for assessing pain in term newborn infants at the bedside. *Clinics (Sao Paulo).* 2012;67(10):1165-70. doi: 10.6061/clinics/2012(10)08.
 24. Giordano V, Edobor J, Deindl P, et al. Pain and Sedation Scales for Neonatal and Pediatric Patients in a Preverbal Stage of Development: A Systematic Review. *JAMA Pediatr.* 2019;173(12):1186-1197. doi: 10.1001/jamapediatrics.2019.3351.
 25. Debillon T, Zupan V, Ravault N, et al. Development and initial validation of the EDIN scale, a new tool for assessing prolonged pain in preterm infants. *Arch Dis Child Fetal Neonatal Ed.* 2001;85(1):F36-41. doi: 10.1136/fn.85.1.f36.
 26. Lundqvist P, Kleberg A, Edberg AK, et al. Development and psychometric properties of the Swedish ALPS-Neo pain and stress assessment scale for newborn infants. *Acta Paediatr.* 2014;103(8):833-9. doi: 10.1111/apa.12672.
 27. Tristão RM, Vilela LR, Marques LA, et al. Adaptation and validation of the EVENDOL scale to assess pain in neonates in Portuguese language. *Early Hum Dev.* 2021;152:105285. doi: 10.1016/j.earlhumdev.2020.105285.
 28. Kappesser J, de Laffolie J, Faas D, et al. Comparison of two neonatal pain assessment tools (Children and Infant's Postoperative Pain Scale and the Neonatal Facial Coding System-Revised) and their relations to clinicians' intuitive pain estimates. *Eur J Pain.* 2018;23(4):708-718. doi: 10.1002/ejp.1338.
 29. Bellieni CV, Cordelli DM, Caliani C, et al. Inter-observer reliability of two pain scales for newborns. *Early Hum Dev.* 2007;83(8):549-52. doi: 10.1016/j.earlhumdev.2006.10.006.

30. Bellieni C, Maffei M, Ancora G, et al. Is the ABC pain scale reliable for premature babies? *Acta Paediatr.* 2007;96(7):1008-10. doi: 10.1111/j.1651-2227.2007.00355.x.
31. Marceau J. Pilot study of a pain assessment tool in the Neonatal Intensive Care Unit. *Journal of paediatrics and child health.* 2003;39(8):598-601. Doi: 10.1046/j.1440-1754.2003.00215.x
32. Milesi C, Cambonie G, Jacquot A, et al. Validation of a neonatal pain scale adapted to the new practices in caring for preterm newborns. *Arch Dis Child Fetal Neonatal Ed.* 2010;95(4):F263-6. doi: 10.1136/adc.2008.144758.
33. Dias FSB, Gasparino RC, Carmona EV, Marba STM. Validation of the Échelle Douleur Inconfort Nouveau-Né for Brazilian culture. *Rev Esc Enferm USP.* 2018;51:e03285. doi: 10.1590/S1980-220X2017008603285.
34. Peters JW, Koot HM, Grunau RE, et al. Neonatal Facial Coding System for assessing postoperative pain in infants: item reduction is valid and feasible. *Clin J Pain.* 2003;19(6):353-63. doi: 10.1097/00002508-200311000-00003.
35. Holsti L, Grunau RE. Initial validation of the Behavioral Indicators of Infant Pain (BIIP). *Pain.* 2007;132(3):264-272. doi: 10.1016/j.pain.2007.01.033.
36. Chen Y, Tong Y, Xue Z, et al. Evaluation of the Reliability and Validity of the Behavioral Indicators of Infant Pain Scale in Chinese Neonates. *Pain Manag Nurs.* 2020;21(5):456-461. doi: 10.1016/j.pmn.2020.01.001.
37. Gibbins S, Stevens BJ, Yamada J, et al. Validation of the Premature Infant Pain Profile-Revised (PIPP-R). *Early Hum Dev.* 2014;90(4):189-93. doi: 10.1016/j.earlhumdev.2014.01.005.
38. Vederhus BJ, Eide GE, Natvig GK. Psychometric testing of a Norwegian version of the Premature Infant Pain Profile: an acute pain assessment tool. A clinical validation study. *Int J Nurs Pract.* 2006;12(6):334-44. doi: 10.1111/j.1440-172X.2006.00592.x.
39. Taplak AŞ, Bayat M. Psychometric Testing of the Turkish Version of the Premature Infant Pain Profile Revised-PIPP-R. *J Pediatr Nurs.* 2019;48:e49-e55. doi: 10.1016/j.pedn.2019.06.007.
40. O'Sullivan AT, Rowley S, Ellis S, et al. The Validity and Clinical Utility of the COVERS Scale and Pain Assessment Tool for Assessing Pain in Neonates Admitted to an Intensive Care Unit. *Clin J Pain.* 2016;32(1):51-7. doi: 10.1097/AJP.0000000000000228.
41. Hummel P, Puchalski M, Creech SD, Weiss MG. Clinical reliability and validity of the N-PASS: neonatal pain, agitation and sedation scale with prolonged pain. *J Perinatol.* 2008;28(1):55-60. doi: 10.1038/sj.jp.7211861.
42. Pölkki T, Korhonen A, Axelin A, et al. Development and preliminary validation of the Neonatal Infant Acute Pain Assessment Scale (NIAPAS). *Int J Nurs Stud.* 2014;51(12):1585-94. doi: 10.1016/j.ijnurstu.2014.04.001.
43. Cignacco E, Mueller R, Hamers JP, Gessler P. Pain assessment in the neonate using the Bernese Pain Scale for Neonates. *Early Hum Dev.* 2004;78(2):125-31. doi: 10.1016/j.earlhumdev.2004.04.001.

44. Sadeghi A, Rassouli M, Gharehdaghi FA, et al. Validation of the Persian Version of Premature Infant Pain Profile-Revised in Hospitalized Infants at the Neonatal Intensive Care Units, Iran J Pediatr. 2017;27(5):e10056. doi: 10.5812/ijp.10056.
45. van Dijk M, Roofthoof DW, Anand KJ, et al. Taking up the challenge of measuring prolonged pain in (premature) neonates: the COMFORTneo scale seems promising. Clin J Pain. 2009;25(7):607-16. doi: 10.1097/AJP.0b013e3181a5b52a.
46. Caljow MAA, Kloos MA, Oliver MY, et al. Measurement of pain in premature infants with a gestational age between 28 to 37 weeks: validation of the adapted COMFORT scale. Journal of Neonatal Nursing. 2007;13(1):13-18. doi: 10.1016/j.jnn.2006.11.007.
47. Liaw JJ, Yang L, Chou HL, et al. Psychometric analysis of a Taiwan-version pain assessment scale for preterm infants. J Clin Nurs. 2012;21(1-2):89-100. doi: 10.1111/j.1365-2702.2011.03770.x.
48. Hummel P, Lawlor-Klean P, Weiss MG. Validity and reliability of the N-PASS assessment tool with acute pain. J Perinatol. 2010;30(7):474-8. doi: 10.1038/jp.2009.185.
49. Huang XZ, Li L, Zhou J, et al. Evaluation of three pain assessment scales used for ventilated neonates. J Clin Nurs. 2018;27(19-20):3522-3529. doi: 10.1111/jocn.14585.
50. Guinsburg R, de Almeida MF, de Araújo Peres C, et al. Reliability of two behavioral tools to assess pain in preterm neonates. Sao Paulo Med J. 2003;121(2):72-6. doi: 10.1590/s1516-31802003000200008.
51. Rivas AL, Rivas ER; Bustos LM. Luis. Validación de escala de valoración del dolor en recién nacidos en una unidad de neonatología. Ciencia y enfermería. 2012;18(2):93-99. doi: 10.4067/S0717-95532012000200010.
52. Woragidpoonpol P, Tiansawad S, Mesukko J, Klunklin P. Development of a Clinical Pain Scale for Preterm Neonates. Pacific Rim International Journal of Nursing Research. 2018;22(4):347-359. doi: https://he02.tci-thaijo.org/index.php/PRIJNR/article/view/95738
53. Malarvizhi G, Vatsa M, Roseline M, et al. Inter-rater reliability of neonatal infant pain scale (NIPS) as a multidimensional behavioral pain tool. Journal of Health and Allied Sciences NU. 2012;2(1):26-30. doi: 10.1055/s-0040-1703551
54. Kappesser J, Kamper-Fuhrmann E, de Laffolie J, et al. Pain-specific Reactions or Indicators of a General Stress Response?: Investigating the Discriminant Validity of 5 Well-established Neonatal Pain Assessment Tools. Clin J Pain. 2019;35(2):101-110. doi: 10.1097/AJP.0000000000000660.
55. Zengin D, Ardahan Sevgili S, Yardimci F, et al. Psychometric properties of the Turkish version of the neonatal infant acute pain assessment scale. J Pediatr Nurs. 2021;(21)00153-6. doi: 10.1016/j.pedn.2021.05.006.
56. Spence K, Gillies D, Harrison D, et al. A reliable pain assessment tool for clinical assessment in the neonatal intensive care unit. J Obstet Gynecol Neonatal Nurs. 2005;34(1):80-6. doi: 10.1177/0884217504272810.
57. Stevens BJ, Gibbins S, Yamada J, et al. The premature infant pain profile-revised

(PIPP-R): initial validation and feasibility. Clin J Pain. 2014;30(3):238-43. doi: 10.1097/AJP.0b013e3182906aed.

Figura 1 - Fluxograma do processo de seleção dos estudos.

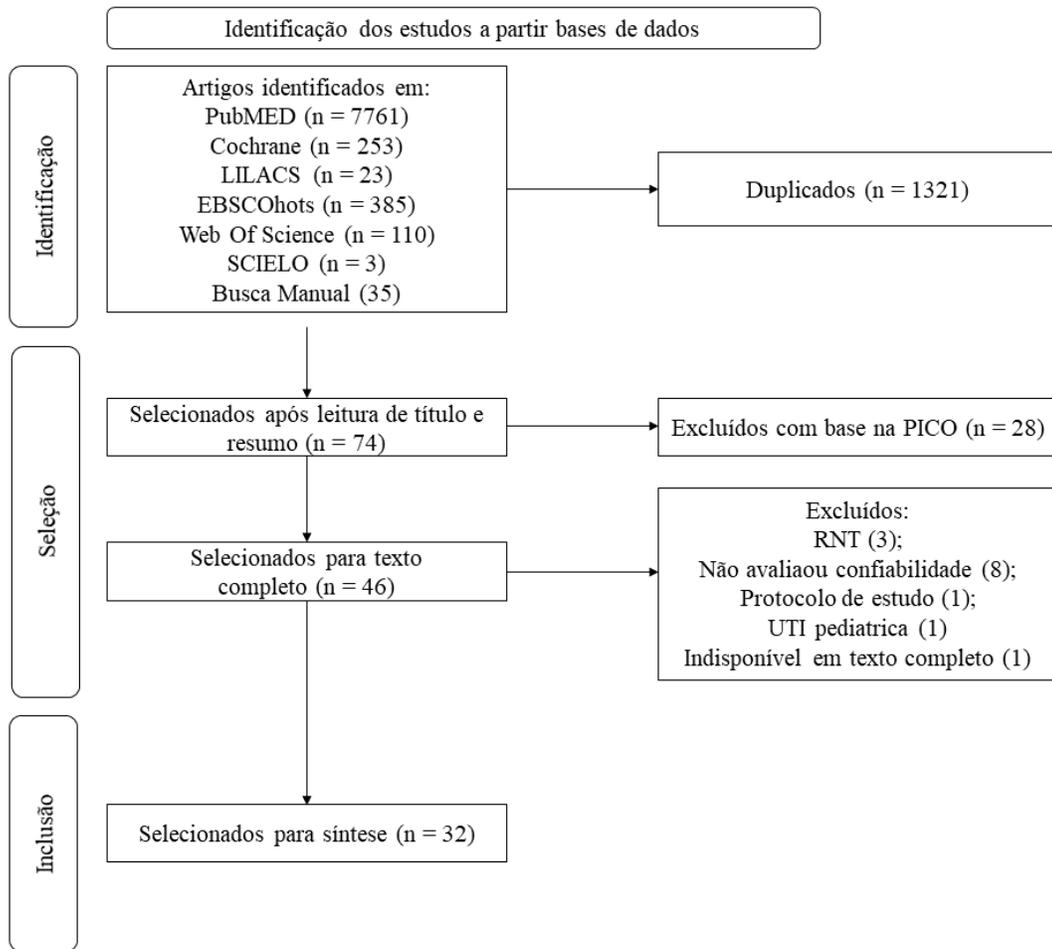


Figura 2 - Gráfico realizado a partir do checklist COSMIN para demonstrar a qualidade dos estudos.

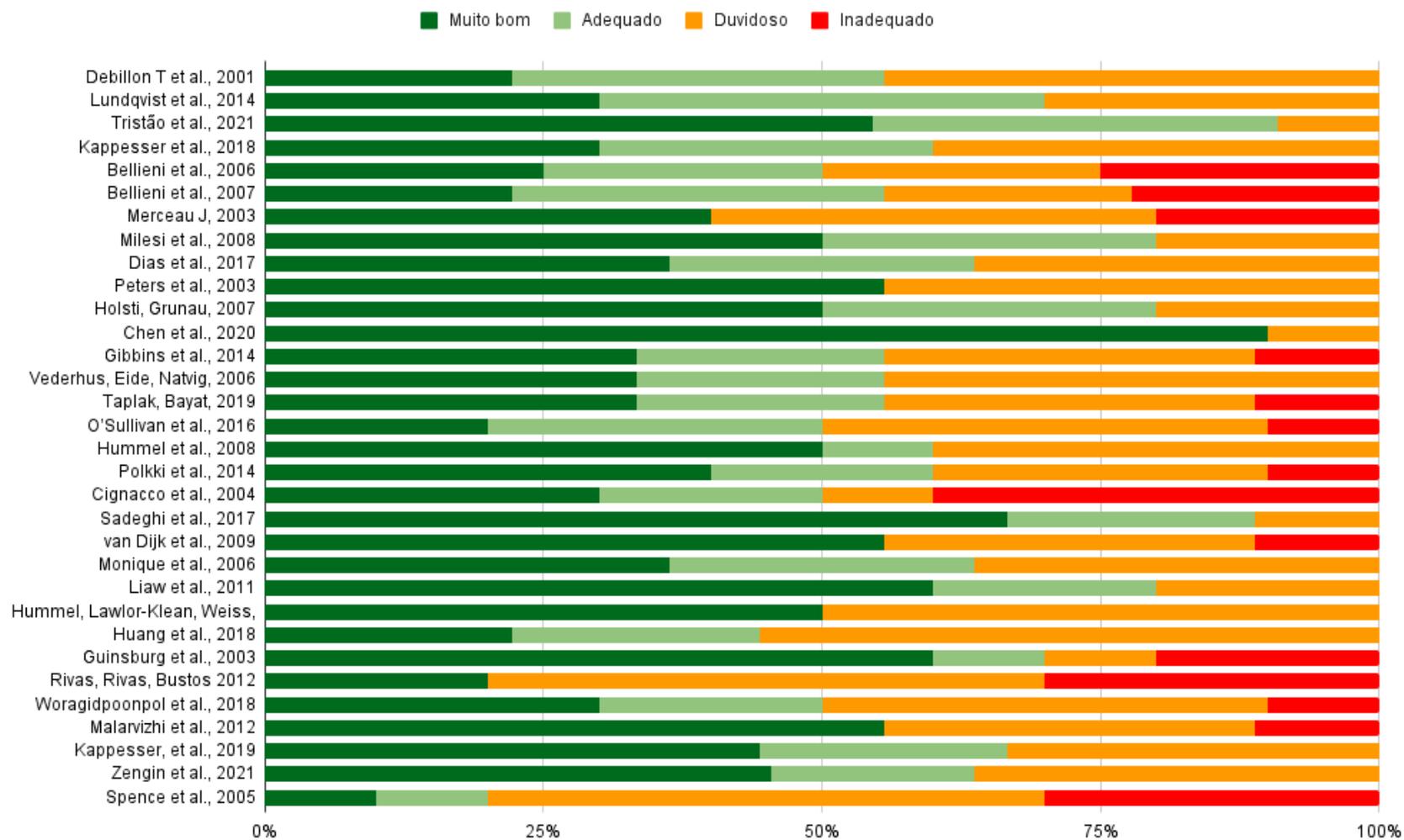


Tabela 1. Descrição da estratégia PICO

Termos para Pesquisa Clínica	Critérios
População	Recém-nascidos pré-termo.
Intervenção/Interesse	Escalas que avaliam comportamento, dor e estresse na UTIN.
Controle	Não foi considerado como critério de seleção
<i>Outcome/Desfecho</i>	Comportamento, estresse e dor

Tabela 2 - Padrões baseados em consenso para a seleção de instrumentos de medição de saúde (COSMIN) - padrões utilizados nesse estudo.

COSMIN						
Nº de Box	Item	Perguntas				
Box 1	1	É fornecida uma descrição clara do construto a ser medido?				
	4	É fornecida uma descrição clara do contexto de uso?				
	13	Para estudos quantitativos (pesquisa): o tamanho da amostra foi adequado?				
Box 4	2	Para pontuações contínuas: o alfa ou ômega de Cronbach foi calculado?				
Box 6	1	Os pacientes estavam estáveis no período intermediário no construto a ser medido?				
	4	Os profissionais administram a medição sem conhecimento dos escores ou valores de outras medições repetidas nos mesmos pacientes?				
	5	Os profissionais atribuíram escores ou determinaram valores sem conhecimento dos escores ou valores de outras medições repetidas nos mesmos pacientes?				
	6	Houve alguma falha importante no projeto ou nos métodos estatísticos do estudo?				
	7	Para pontuações contínuas: foi calculado o coeficiente de correlação intraclasse (ICC)?				
	8	Para pontuações ordinais: um kappa ponderado foi calculado?				
	9	Para pontuações dicotômica/nominal: foi calculado o Kappa para cada categoria contra outras categorias combinadas?				
Box 9	1	Está claro o que o (s) instrumento (s) comparador (es) mede (m)?				
Classificação		Muito bom	Adequado	Duvidoso	Inadequado	Não se aplica

Tabela 3 - Descrição e separação das escalas encontradas

Escalas	
Unidimensionais	Multidimensionais
<i>ABC pain scale</i>	<i>BPSN - Bernese Pain Scale for Neonates</i>
<i>EVENDOL behavioral pain scale</i>	<i>NIPS - Neonatal Infant Pain Scale</i>
<i>BIIP - Behavioral Indicators of Infant Pain</i>	<i>PASPI - Pain Assessment Scale for Preterm Infants</i>
<i>NFCS - Neonatal Facial Coding System</i>	<i>NNICUPAT - Nepean Neonatal Intensive Care Unit Pain Assessment Tool</i>
<i>NFCS-R - Neonatal Facial Coding System - Revised</i>	<i>N-PASS - Neonatal Pain Agitation and Sedation Scale</i>
<i>COMFORT-neo</i>	<i>PAT - Pain Assessment Tool</i>
<i>ALPS-neo - Astrid Lindgren and Lund Children's Hospitals Pain and Stress Assessment Scale for Preterm and sick Newborn Infants</i>	<i>COMFORT adaptada</i>
<i>CHIPPS - Children's and Infants Postoperative Pain Scale</i>	<i>PIPP - Premature Infant Pain Profile</i>
<i>CSS - Clinical Scoring System</i>	<i>PIPP-R - Premature Infant Pain Profile - Revised</i>
<i>GIVES Neonatal Pain Assessment Scale</i>	<i>FANS - Faceless Acute Neonatal Pain Scale</i>
<i>EDIN - Échelle Douleur Inconfort Nouveau-Né</i>	<i>COVERS neonatal pain scale</i>
	<i>NIAPAS - Neonatal Infant Acute Pain Assessment Scale</i>
	<i>CPSPN - Clinical Pain Scale for Preterm Neonates</i>

Bellieni et al., 2007	ABC	Dor aguda	-	0.70	-	-	-	-	-	-	-	-
Merceau J, 2003	NNICUPAT	Dor aguda em VM	0.48-0.88	-	-	-	-	-	-	-	-	-
Milesi et al., 2010	FANS	Dor aguda	0.92	-	-	-	-	-	-	0.72	-	-
Holsti, Grunau, 2007	BIIP	Comportamento	Pré - 0.80 Durante - 0.92 Pós - 0.88	-	-	-	-	-	-	0.82	-	-
Chen et al., 2020	BIIP	Comportamento	0.92-0.95	-	-	-	-	-	-	0.90	0.94	-
Malarvizhi et al., 2012	NIPS	Dor aguda	-	-	-	0.71-0.96	-	-	-	0.90	-	-
Bellieni et al., 2006	NIPS / PIPP	Dor aguda	0.59 / 0.16	0.59-0.60 / 0.10-0.15	-	-	-	-	-	-	-	-
Vederhus, Eide, Natvig, 2006	PIPP	Dor aguda	-	-	-	0.89-0.97	-	-	-	0.78	-	-
Gibbins et al., 2014	PIPP-R	Dor aguda	-	-	-	0.87-0.92	-	-	-	-	-	-
Taplak, Bayat, 2019	PIPP-R	Dor aguda	0.94-1.00	-	-	-	-	-	-	0.84	-	91.50
Sadeghi et al., 2017	PIPP-R	Dor aguda	0.98-0.99	-	-	-	-	-	-	0.71	-	-
Huang et al., 2018 (Xiao-Zhi)	PIPP-R/N-P ASS/ NIAPAS	Dor aguda	0.99	-	-	-	-	-	-	0.73 / 0.83 / 0.83	-	-
Hummel et al., 2008	N-PASS	Dor prolongada / Sedação	Dor: Pré - 0.97 Pós - 0.96 / Sedação:	-	-	-	-	-	-	Dor: Obs 1 - 0.82 Obs 2 - 0.72 Sedação: Obs 1 - 0.89	-	-

			Pré - 0,92 Pós - 0,95							Obs 2 - 0,89		
Hummel, Lawlor-Klean, Weiss, 2010	N-PASS	Dor aguda	0,93	-	-	-	-	-	-	0,84-0,89	-	-
Pölkki et al., 2014	NIAPAS	Dor aguda	-	-	-	0,99	-	0,99-1,00	-	0,72	-	-
Zengin et al., 2021	NIAPAS	Dor aguda	Pré - 0,95 Durante - 0,93 Pós - 0,92	-	0,66-0,92	-	-	-	-	0,95-0,97	Obs 1 0,93-0,97 Obs 2 0,91-0,98	-
O'Sullivan et al., 2016	COVERS / PAT	Dor aguda	0,80-0,82 / 0,83-0,86	0,39-0,44/ 0,34-0,41	-	-	-	-	-	0,74-0,78/ 0,79-0,85	-	-
van Dijk et al., 2009	COMFORT- neo	Dor prolongada	-	-	0,79	-	-	-	-	0,84-0,88	-	81,00
Monique et al., 2006	COMFORT adaptada	Dor aguda	0,84-0,93	-	0,29-0,84	-	-	-	-	0,76-0,86	-	-
Liaw et al., 2011	PASPI	Dor aguda	0,88-0,93	-	-	-	-	-	-	0,84	-	-
Rivas, Rivas, Bustos 2012	GIVES	Não relata	-	-	-	-	-	-	-	0,78	-	-
Woragidpoonpol et al., 2018	CPSPN	Dor aguda	0,91-1,00	-	-	-	-	-	-	0,94	-	-
Spence et al., 2005	PAT	Dor aguda	0,85	-	-	-	-	-	-	-	-	-

Abreviações: EDIN, Échelle Douleur Inconfort Nouveau-Né; ALPS-neo, Astrid Lindgren and Lund Children's Hospitals Pain and Stress Assessment Scale for Preterm and sick Newborn Infants; EVENDOL, behavioral pain scale; NFCS-R, Neonatal Facial Coding System - Revised; CHIPPS, Children's and Infants Postoperative Pain Scale; PIPP-R, Premature Infant Pain Profile - Revised; NIPS, Neonatal Infant Pain Scale; N-PASS, Neonatal Pain Agitation and Sedation Scale; BPSN, Bernese Pain Scale for Neonates; NFCS, Neonatal Facial Coding System; CSS, Clinical Scoring System; ABC, pain scale; NNICUPAT, Nepean Neonatal Intensive Care Unit Pain Assessment Tool; FANS, Faceless Acute Neonatal Pain Scale; BIIP, Behavioral Indicators of Infant Pain; PIPP, Premature Infant Pain Profile; NIAPAS, Neonatal Infant Acute Pain Assessment Scale; COVERES, neonatal pain scale; PAT, Pain Assessment Tool; PASPI, Pain Assessment Scale for Preterm Infants; GIVES, Neonatal Pain Assessment Scale; CPSPN, Clinical Pain Scale for Preterm Neonates; ICC, Coeficiente de Correlação Intraclasse ; Obs 1, Observador 1; Obs 2, Observador 2; Pré/durante/pós, fases da intervenção realizada.

Tabela 5 - Descrição das Escalas

Scales	Indicators	Score	Description	Total Score
<i>Behavioral Indicators of Infant Pain - BIIP</i>	Sleep/Wake States			0-9 points 0-2: comfortable 3-6: moderate pain 7 a 9: severe pain
	Deep Sleep	0	Eyes closed, regular breathing, no movements of extremities.	
	Active Sleep	0	Eyes closed, twitches or startles of extremities, rapid eye movements, irregular breathing.	
	Drowsy	0	Eyes open (but roving or not focused) or closed, irregular breathing, some body movements.	
	Quiet Awake	0	Eyes open, focused, very few or no body movements.	
	Active Awake	1	Eyes open, active extremity movements.	
	Agitated/Crying	2	Upset, fussing, highly aroused, crying.	
	Face & Hand Actions			
	Brow bulge	1	Bulging, creasing and/or vertical furrows above and between brows occurring as a result of lowering and drawing together of the eyebrows.	
	Eye squeeze	1	Squeezing and/or bulging of the eyelids.	
	Naso-labial furrow	1	Pulling upwards and deepening of the naso-labial furrow (a line or wrinkle which begins adjacent to the nostril wings and runs down and outwards beyond the lip corners).	
	Horizontal mouth stretch	1	A distinct horizontal stretch pull at the corners of the mouth sometimes accompanied by a taut upper lip.	
	Taut tongue	1	Raised, cupped tongue with sharp tensed edges. The first occurrence of taut tongue is usually easy to see, often occurring with a wide open mouth. After this first occurrence, the mouth may close slightly. Taut tongue can be scored on the basis of the still visible tongue edges.	

	Finger splay	1	Sudden opening of the hands with fingers extended and separated from each other.		
	Fisting	1	Tight closing and flexing of the fingers to form a fist.		
<i>Premature Infant Pain Profile Revised - PIPP-R</i>	Change in Heart Rate	0, +1, +2, +3	0 = 0-4; +1 = 5-14; +2 = 15-24 e +3 = >24	Score calculation: Total + Gestational Age + Alert State 0-18 points to Term Infants 0-21 points to Preterm Infants <6: mild pain 7-12: moderate pain >13: severe pain	
	Decrease in Oxygen Saturation	0, +1, +2, +3	0 = 0-2; +1 = 3-3; +2 = 6-8 e +3 = >8 ou Increase in O ₂		
	Brow Bulge (sec)	0, +1, +2, +3	0 = none (<3); +1 = minimal (3-10); +2 = moderate (11-20) e +3 = maximal (>20)		
	Eye Squeeze (sec)	0, +1, +2, +3	0 = none (<3); +1 = minimal (3-10); +2 = moderate (11-20) e +3 = maximal (>20)		
	Naso-Labial furrow (sec)	0, +1, +2, +3	0 = none (<3); +1 = minimal (3-10); +2 = moderate (11-20) e +3 = maximal (>20)		
	*Sub-Total Score				
	Gestational Age (Wks + Days)	0, +1, +2, +3	0 = ≥36 wks; +1 = 32-35 wks, 6d; +2 = 28-31 wks, 6d e +3 = <28 wks		
	Baseline Behavioural State	0, +1, +2, +3	0 = active and awake; +1 = quiet and awake; +2 = active and asleep +3 = quiet and asleep		
	** Total score				
<i>Evaluation Enfant Douleur - EVENDOL</i>	Vocal or verbal expression	0, 1, 2, 3	0 = sign obsent; 1 = sign weak or transient; 2 = sign moderate or present about half the time; 3 = sign strong or present almost all the time	0-15 points Limit value to treatment 4/15	
	Facial expression	0, 1, 2, 3	0 = sign obsent; 1 = sign weak or transient; 2 = sign moderate or present about half the time; 3 = sign strong or present almost all the time		
	Movements	0, 1, 2, 3	0 = sign obsent; 1 = sign weak or transient; 2 = sign moderate or present about half the time; 3 = sign strong or present almost all the time		

	Postures	0, 1, 2, 3	0 = sign absent; 1 = sign weak or transient; 2 = sign moderate or present about half the time; 3 = sign strong or present almost all the time	
	Interaction with the environment	0, 1, 2, 3	0 = normal; 1 = low; 2 = very low; 3 = absent	
<i>Pain Assessment Scale for Preterm Infants - PASPI</i>	Transition between states	0, 1, 2, 3	States: quiet sleep; active sleep; transition from sleep to awake; quiet awake; active awake and crying 0 = maintaining the same state; 1 = change from sleep to drowsy or from drowsy to awake; 2 = change from sleep or drowsy to awake with fussing; 3 = change from sleep, drowsy or awake to loudly crying	0-18 points Higher scores indicate more severe pain
	Facial expression	0, 1, 2, 3	Brow bulge, eye squeeze, nasolabial furrow and mouth stretch; 0 = eyebrow and face relaxed; 1 = slight brow bulge and eye squeeze; 2 = moderate brow bulge and eye squeeze, nasolabial furrow and mouth stretch; 3 = severe brow bulge and eye squeeze, nasolabial furrow and mouth stretch	
	HR	0, 1, 2, 3	0 = normal and regular; 1 = increased or decreased 12-16 beats from baseline mean; 2 = increased or decreased 17-24 beats from baseline mean; 3 = increased or decreased ≥ 25 beats from baseline mean	
	Oxygen saturation	0, 1, 2, 3	0 = decreased 2.0% from baseline mean; 1 = decreased 3.0-5.0% from baseline mean; 2 = decreased 6.0-7.0% from baseline mean; 3 = decreased 8.0% from baseline mean	
	Limb, body movement	0, 1, 2, 3	0 = no movements in between 10 seconds; 1 = movements in the first interval of 10 seconds; 2 = continuation of the movements in the second interval; 3 = continuation of the movements in the third interval.	
	Hand behaviour	0, 1, 2, 3	0 = no movements in between 10 seconds; 1 = movements in the first interval of 10 seconds; 2 = continuation of the movements in the second interval; 3 = continuation of the movements in the third interval	

ANEXO 1 - REGRAS DA REVISTA

Title of Manuscript [97 character max, including spaces]

Alice B. Smith^a, MD, Clarence Jones^{a,b}, MD, PhD, Ann S. Noble^b, MPH, Steve R. Eccleston^b, MPH, William E. Piper^a, PhD, Ellen Johnson, MSc, Murray Green, PhD, on behalf of Authoring Group (if any)

Affiliations: ^aChildren's Hospital, Chicago, Illinois; and ^bUniversity of Chicago, Chicago, Illinois

Address correspondence to: Alice Smith, Department of Pediatrics, Children's Hospital, 1234 Main Street, Chicago IL, 60641, [asmith@example.com], 773-900-9000.

Short title: Short running title for Manuscript [55 characters maximum, including spaces]

Conflict of Interest Disclosures (includes financial disclosures): Ann Noble has example conflict. Clarence Jones has other example conflict. The other authors have no example conflicts of interest to disclose.

Funding/Support: All phases of this study were supported by an example NIH grant, #####. [or] No funding was secured for this study.

Role of Funder/Sponsor (if any): The NIH example had no role in the design and conduct of the study. Bayer Healthcare example provided reagents for B-type natriuretic peptide testing.

Clinical Trial Registration (if any): Registry name, registration number, web link to study on registry, and data sharing statement

Abbreviations: (list and define abbreviations used in the text; [or] Abbreviations: none)

Article Summary

Max 25 words; a brief insight into what the article is about; it should entice the reader to go beyond the table of contents page and read the full article.

Contributors' Statement Page

[Each author should only appear once. Use names, not initials. If multiple authors have identical contributions, list them in the same sentence; otherwise, list separately.]

Dr Smith and Prof Jones conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript.

Drs Noble, Eccleston, and Piper and Ms Johnson designed the data collection instruments, collected data, carried out the initial analyses, and reviewed and revised the manuscript.

Dr Green conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Abstract [250 words max]

Context: Context content goes here. Context content goes here. Context content goes. Context content goes

Objective: Objective content goes here. Objective content goes here. Objective content goes here.

Data Sources: Data sources content goes here. Data sources content goes here. Data sources content goes here.

Study Selection: Study selection content goes here. Study selection content goes here. Study selection content goes here.

Data Extraction: Data extraction content goes here. Data extraction content goes here. Data extraction content goes here.

Results: Results content goes here. Results content goes here. Results content goes here.

Limitations: Limitations content goes here. Limitations content goes here. Limitations content goes here.

Conclusions: Conclusions content goes here. Conclusions content goes here. Conclusions content goes here.

[The Main Body of the Manuscript Goes Here; Review = 4,000 words max]

This is my manuscript's main body. It's in double-spaced, Times New Roman font, size 12.

For a REVIEW article, this section can be up to 4,000 words long.¹ This is my manuscript's main body.² It's in double-spaced, Times New Roman font, size 12. For a REVIEW article, this section can be up to 4,000 words long.³

This is my manuscript's main body. It's in double-spaced, Times New Roman font, size 12.

For a REVIEW article, this section can be up to 4,000 words long.¹ This is my manuscript's main body.² It's in double-spaced, Times New Roman font, size 12. For a REVIEW article, this section can be up to 4,000 words long.³

This is my manuscript's main body. It's in double-spaced, Times New Roman font, size 12.

For a REVIEW article, this section can be up to 4,000 words long.¹ This is my manuscript's main body.² It's in double-spaced, Times New Roman font, size 12. For a REVIEW article, this section can be up to 4,000 words long.³

This is my manuscript's main body. It's in double-spaced, Times New Roman font, size 12.

For a REVIEW article, this section can be up to 4,000 words long.¹ This is my manuscript's main body.² It's in double-spaced, Times New Roman font, size 12. For a REVIEW article, this section can be up to 4,000 words long.³

Acknowledgments [Optional]

Thank you to Dr. Jane Doe for her review of the manuscript. Thank you also to the research group at College University Medical Center for their contribution

References

1. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
2. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
3. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
4. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
5. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
6. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
7. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
8. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
9. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.
10. Smith A, Jones B, King G, et al. Article title goes here, 1900-2012. Pediatrics. 2014;123(1):e75-e82. doi: 10.1542/peds.2015-1827.

Figure 1. This is the legend from a figure from the study.

Figure 2. This is also a legend from a figure from the study.

[Each figure must include a legend, listed in numerical order at the end of your manuscript. Do not include figure legends in the figures themselves. Number figures in the order in which they appear in the text. Upload figures as individual files; see the author guidelines for image types and further formatting requirements.]

[Tables may appear at the end of your main document or be uploaded separately.]

[We cannot accept Excel (.xls, .xlsx) or Powerpoint (.ppt, .pptx) files.]