



**MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CAMPUS MINISTRO REIS VELLOSO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOMÉDICAS –
PPGCBM**

BRUNA DE MORAES RUBIM ALELAF

**INSTRUMENTOS DE AVALIAÇÃO DA DOR CRÔNICA EM CRIANÇAS
HOSPITALIZADAS: UMA REVISÃO SISTEMÁTICA DAS
PROPRIEDADES DE MEDIDA**

**PARNAÍBA – PI
NOVEMBRO/ 2019**

BRUNA DE MORAES RUBIM ALELAF

**INSTRUMENTOS DE AVALIAÇÃO DA DOR CRÔNICA EM CRIANÇAS
HOSPITALIZADAS: UMA REVISÃO SISTEMÁTICA DAS
PROPRIEDADES DE MEDIDA**

Dissertação apresentada ao Programa de Pós-graduação em Ciências Biomédicas da Universidade Federal do Piauí, *Campus* Ministro Reis Velloso, como requisito para a obtenção do título de Mestre em Ciências Biomédicas.

Área de concentração: Desempenho humano: diagnóstico e funcionalidade.

Orientador: Prof. Dr. Fuad Ahmad Hazime.

**PARNAÍBA – PI
NOVEMBRO/ 2019**

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Setorial Prof. Cândido Athayde – Campus Parnaíba
Serviço de Processamento Técnico

A366i Alelaf, Bruna de Moraes Rubim

Instrumentos de avaliação da dor crônica em crianças hospitalizadas:
uma revisão sistemática das propriedades de medida [recurso eletrônico] /
Bruna de Moraes Rubim Alelaf. – 2019.

1 Arquivo em PDF

Dissertação (Mestrado em Ciências Biomédicas) - Universidade Federal
do Piauí, 2019.

Orientação: Prof. Dr. Fuad Ahmad Hazime

1. Avaliação. 2. Crianças. 3. Dor Crônica. 4. Pacientes Internados. I.
Titulo.

CDD: 616.0472

BRUNA DE MORAES RUBIM ALELAF

**INSTRUMENTOS DE AVALIAÇÃO DA DOR CRÔNICA EM CRIANÇAS
HOSPITALIZADAS: UMA REVISÃO SISTEMÁTICA DAS
PROPRIEDADES DE MEDIDA**

Dissertação apresentada ao Programa de Pós-graduação em Ciências Biomédicas da Universidade Federal do Piauí, *Campus* Ministro Reis Velloso, como requisito para a obtenção do título de Mestre em Ciências Biomédicas.

APROVADA EM: ___/___/___

BANCA EXAMINADORA

Prof. Dr. Baldomero Antonio Kato da Silva
Universidade Federal do Piauí – UFPI

Prof. Dr. Cristiano Sales da Silva
Universidade Federal do Piauí – UFPI

Prof. Dr. Fuad Ahmad Hazime
Universidade Federal do Piauí – UFPI
(Orientador)

**PARNAÍBA – PI
NOVEMBRO/ 2019**

AGRADECIMENTOS

Ao meu esposo, por sempre me apoiar nas minhas decisões, pela paciência, companheirismo, por me acalmar sempre que necessário, por me mostrar o lado bom das coisas quando eu acho que não vai dar certo e por me incentivar na minha vida acadêmica e profissional. Muito obrigada por ser este esposo tão participativo e amável.

Aos meus pais, pelo apoio e torcida incondicional, por me mostrarem o melhor caminho, pela educação e por estarem sempre ao meu lado, em todos os momentos, mesmo que distantes. Não sei o que seria de mim sem vocês.

À minha irmã, que sempre está ao meu lado, me apoiando, cuidando, mostrando o melhor caminho e chamando minha atenção, quando necessário. Eu não seria a mesma sem as suas orientações, que me tornaram uma pessoa melhor.

Ao meu cunhado, Marcelo, pelo incentivo ao caminho da docência e realização de um sonho, que é o mestrado.

Ao meu orientador, Fuad Ahmad Hazime, pela sua disponibilidade, dedicação, paciência, incentivo, ensino, confiança e colaboração em todo este caminho. Por ser tão paciente e calmo, até nos momentos de chamar atenção. Por me orientar no conhecimento sobre a dor crônica, que se tornou tão significativa na minha vida profissional e acadêmica.

Ao Programa de Pós-graduação em Ciências Biomédicas da Universidade Federal do Piauí e todo o corpo docente, na pessoa do coordenador Prof. Dr. Baldomero Antonio Kato da Silva.

Aos membros da banca nos processos de qualificação e defesa, Prof. Dr. Baldomero Antonio Kato da Silva, Prof^a. Dra. Alessandra Tanuri Magalhães, Prof. Dr. Cristiano Sales da Silva, pela atenção dispensada e preciosas sugestões.

Aos professores, Dr. Emerson Diogenes de Medeiros, Dr. Abrahão Fontes Baptista e Dra. Katia Nunes Sá, pela contribuição significativa na realização da pesquisa.

Aos alunos do laboratório LANDS, que me receberam de braços abertos, me apoiaram e incentivaram, mostrando que uma equipe unida rende muito aprendizado. Tenho muito orgulho de fazer parte desta equipe super dedicada e especial.

Aos meus colegas de trabalho e minha coordenadora, Karlla Leite, pelo apoio nos momentos que precisei me ausentar do trabalho, devido as viagens para a realização do mestrado.

Às crianças que atendo na unidade em que trabalho, sem as quais eu não teria me tornado uma profissional de enfermagem preocupada e envolvida com a pesquisa, com a intenção de melhorar suas condições de saúde, priorizando a qualidade de vida.

À todos que, direta ou indiretamente, participaram desta caminhada e contribuíram para a sua concretização. Muito obrigada!

RESUMO

Introdução: A avaliação da dor crônica em crianças hospitalizadas é uma tarefa extremamente desafiadora para cuidadores, pesquisadores e profissionais da saúde. Além das dificuldades inerentes à falta de maturação neuropsicomotora, estas crianças podem ainda apresentar alterações cognitivas, limitando a interpretação e confiabilidade dos instrumentos de avaliação da dor. A imprecisão destes instrumentos dificulta o manejo eficaz da dor, uma vez que podem não ser confiáveis para medir o construto que se pretende. **Objetivo:** Sintetizar e analisar sistematicamente a qualidade das propriedades de medida de instrumentos aplicados na avaliação de dor crônica em crianças hospitalizadas. **Metodologia:** Revisão sistemática conduzida de acordo com as recomendações do *Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA)*. Foram incluídos estudos que abordassem instrumentos de avaliação da dor crônica em crianças, ≥ 2 anos e < 12 anos, hospitalizadas, e que apresentassem em seus resultados a avaliação de, no mínimo, uma propriedade de medida. As bases de dados Pubmed, Web of Science, Cochrane, Cinahl, PsycNET, Embase e Google Scholar foram pesquisadas por revisores independentes. A qualidade dos estudos incluídos foi avaliada por meio do (1) *checklist* Risco de Viés do *CO*n*SENSUS*-based *ST*andards for the selection of health status *ME*asurement *IN*struments (COSMIN), (2) critérios para boas propriedades de medição e abordagem GRADE do COSMIN. A extração dos dados foi realizada por meio de uma ferramenta padronizada do Instituto Joanna Briggs (JBI) e adaptada aos elementos específicos de uma revisão psicométrica. **Resultados:** Foram incluídos na revisão dois estudos, os quais avaliaram o Questionário de Experiência da Dor para Crianças (PEQ-C) e a Avaliação Hétero Douleur Enfant (HEDEN). As propriedades de medida avaliadas foram a validade estrutural, consistência interna, confiabilidade, validade de critério e validade de construto. A maioria das propriedades apresentaram classificação muito boa, conforme *checklist* Risco de Viés e classificação insuficiente, segundo os critérios para boas propriedades de medida. Já as evidências foram classificadas, segundo a abordagem GRADE, como moderada na maioria das avaliações. **Conclusão:** Apesar da existência de uma grande quantidade de instrumentos para avaliação da dor, poucos são direcionados especificamente para a dor crônica em crianças hospitalizadas. Neste sentido, urge a necessidade de mais estudos sobre o tema, seguindo as recomendações de validação e avaliação das propriedades psicométricas reconhecidas.

Palavras-chave: Avaliação, Crianças, Dor crônica, Pacientes internados.

ABSTRACT

Introduction: The evaluation of chronic pain in hospitalized children is an extremely challenging task for caregivers, researchers and health professionals. In addition to the difficulties inherent in the lack of neuropsychomotor maturation, these children may still present cognitive alterations, limiting the interpretation and reliability of pain assessment instruments. The imprecision of these instruments affect the efficacy of pain management, since they may not be reliable to measure the construct that is intended. **Objective:** Systematically synthesize and analyze the quality of the properties of instrument measures applied in the evaluation of chronic pain in hospitalized children. **Methodology:** Systematic review conducted according to the recommendations of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). We included studies that addressed instruments for chronic pain evaluation in hospitalized children ≥ 2 years and <12 years, and that presented in their results the evaluation of at least one measurement property. The Pubmed, Web of Science, Cochrane, Cinahl, PsycNET, Embase, and Google Scholar databases were searched by independent reviewers. The quality of included studies was assessed using the (1) COnsensus-based Bias Risk checklist (COSMIN), (2) criteria for good measurement properties, and COSMIN's GRADE approach. Data extraction was performed using a standardized tool from the Joanna Briggs Institute (JBI) and adapted to the specific elements of a psychometric review. **Results:** Two studies were included in the review, which evaluated the Pain Experience Questionnaire for Children (PEQ-C) and the Hétero Evaluation Douleur Enfant (HEDEN). The measurement properties evaluated were structural validity, internal consistency, reliability, criterion validity and construct validity. Most properties presented very good classification, according to the Risk of Bias checklist and insufficient classification, according to the criteria for good measurement properties. Evidence was classified according to the GRADE approach as moderate in most evaluations. **Conclusion:** Despite the existence of a large number of pain assessment instruments, few are specifically directed to chronic pain in hospitalized children. In this sense, there is a need for further studies on the subject, following the recommendations for validation and evaluation of recognized psychometric properties.

Keywords: Evaluation, Children, Chronic pain, Inpatients.

LISTA DE TABELAS

Tabela 01: Instrumentos de avaliação da dor encontrados em estudos excluídos após leitura do artigo na íntegra.....	29
Tabela 02: Análise de qualidade metodológica nos estudos através do <i>checklist</i> Risco de Viés COSMIN	38
Tabela 03: Classificação de acordo com os critérios para boas propriedades de medição.....	39
Tabela 04: Tabela Resumo da Descoberta (<i>Summary of Finding</i> - SoF) e avaliação GRADE	40
Tabela 05: Dados extraídos dos estudos classificados para a revisão sistemática..	42

LISTA DE ABREVIATURAS

CAS: Escala Analógica de Cor (*Color Analogue Scale*)

CAT-II: Teste de Ansiedade Infantil II (*Children's Anxiety Test-II*)

ICC: Coeficiente de Correlação Intraclasse (*Intraclass Correlation Coefficient*)

CFA: Análise Fatorial Confirmatória (*Confirmatory Factor Analysis*)

CFI: Índice de Ajuste Comparativo (*Comparative Fit Index*)

COSMIN: Padrões baseados em COnsensus para a seleção de INstrumentos de Medição de saúde (*COnsensus-based Standards for the selection of health Measurement INstruments*)

CSI: Inventário de Somatização Infantil (*Children's Somatization Inventory*)

DEGR: Dor da Criança Gustave Roussy (*Douleur Enfant Gustave Roussy*)

DTC: Teste de Depressão para Crianças (*Depression Test for Children*)

EVA: Escala Visual Analógica (*Visual Analogue Scale*)

FLACC: Escala Face, Pernas, Atividade, Choro e Consolabilidade (*Face, Legs, Activity, Cry and Consolability Scale*)

GFI: Índice de Qualidade do Ajuste (*Goodness of Fit Index*)

HEDEN: Avaliação Hétero *Douleur Enfant*

IVC: Índice de Validade de Conteúdo

FPS-R: Escala de Dor Facial – Revisada (*Faces Pain Scale – Revised*)

JBISRIR: Banco de Dados Instituto Joanna Briggs de Revisões Sistemáticas e Relatórios de Implementação (*Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports*)

JBI SUMARI: Sistema para o Gerenciamento Unificado, Avaliação e Revisão de Informações do Instituto Joanna Briggs (*The system for unified management, evaluation and review of information from the Joanna Briggs Institute*)

MPI: Inventário de Dor Multidimensional (*Multidimensional Pain Inventory*)

MPPQ: Questionário de Práticas de Parentalidade Materna (*Maternal Parenting Practice Questionnaire*)

NRS - 11: Escala Numérica de Avaliação – 11 (*Numerical Rating Scale - 11*)

OMS: Organização Mundial de Saúde

PCI: Inventário dos Filhos Parentais (*Parental Childrearing Inventory*)

PEQ-C: Questionário de Experiência da Dor para crianças (*Pain Experience Questionnaire for Children*)

PRCQ-C: Questionário de Cognições Relacionadas à Dor para Crianças e Adolescentes (*Pain-Related Cognitions Questionnaire for children and adolescents*)

PRISMA: Itens de Relatórios Preferidos para Revisões Sistemáticas e Meta-Análises (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*)

PROSPERO: Registro Internacional Prospectivo de Revisões Sistemáticas (*International Prospective Register of Systematic Reviews*)

RMSA: Erro Quadrático Médio da Raiz da Aproximação (*Root Mean Square Error of Approximation*)

SDQ: Questionário de Pontos Fortes e Dificuldades (*Strengths and Difficulties Questionnaire*)

SoF: Resumo da Descoberta (*Summary of Finding*)

SUMÁRIO

CONSIDERAÇÕES INICIAIS.....	12
1. INTRODUÇÃO	13
2. REFERENCIAL TEÓRICO.....	15
2.1 Dor em crianças	15
2.2 Dor crônica	16
2.3 Hospitalização	17
2.4 Instrumentos de avaliação	18
2.5 Propriedades de medida.....	20
3. OBJETIVOS.....	23
3.1 Objetivo geral	23
3.2 Objetivos específicos.....	23
4. METODOLOGIA	24
4.1 Critérios de inclusão	24
4.1.1 Participantes.....	24
4.1.2 Construto.....	24
4.1.3 Instrumentos.....	24
4.1.4 Resultados	24
4.1.5 Tipos de estudo.....	25
4.2 Estratégia de busca.....	25
4.3 Seleção de estudos	26
4.4 Avaliação da qualidade metodológica.....	26
4.5 Extração e síntese de dados	27
5. RESULTADOS.....	28

5.1 Análise da qualidade metodológica nos estudos através do <i>checklist</i> Risco de Viés COSMIN	31
5.2 Avaliação de acordo com os critérios para boas propriedades de medição	31
5.3 Avaliação das evidências sumarizadas e classificação conforme abordagem GRADE	32
5.4 Dados extraídos dos estudos incluídos na revisão	34
5.4.1 Características dos estudos incluídos	34
5.4.2 Características da amostra de inclusão.....	34
5.4.3 Características das ferramentas de avaliação da dor crônica	35
5.4.4 Propriedades de medição.....	35
5.4.5 Descrição dos resultados	35
6. DISCUSSÃO	45
7. CONSIDERAÇÕES FINAIS	51
REFERÊNCIAS.....	52
APÊNDICE I.....	65
APÊNDICE II.....	72
APÊNDICE III.....	73
ANEXO I.....	75
ANEXO II.....	136
ANEXO III.....	139

CONSIDERAÇÕES INICIAIS

O presente documento é parte dos requisitos para a realização da Defesa no curso de Mestrado do Programa de Pós-graduação em Ciências Biomédicas da Universidade Federal do Piauí.

Este documento está organizado em duas partes, a saber:

I. Revisão sistemática - Instrumentos de avaliação da dor crônica em crianças hospitalizadas: uma revisão sistemática das propriedades de medida.

II. Manuscrito submetido no periódico ***Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports (JBISRIR)***, que discorre sobre o protocolo de revisão sistemática das propriedades de medidas de instrumentos de avaliação da dor crônica em crianças hospitalizadas.

1. INTRODUÇÃO

A dor crônica pediátrica é uma condição bastante desafiadora que afeta aproximadamente 20% a 35% das crianças em todo o mundo (Goodman e Mcgrath, 1991; Stanford *et al.*, 2008; King *et al.*, 2011), variando conforme a origem primária, como cefaleia (8-83%), dor abdominal (4-53%) ou musculoesquelética (4-40%) (King *et al.*, 2011). É definida como dor contínua ou recorrente que dura por 3 meses ou mais, podendo ser incapacitante (Huguet e Miro, 2008; King *et al.*, 2011). A persistência da dor na infância pode ainda estar presente na idade adulta (Hassett *et al.*, 2013) e impactar negativamente na sua condição física e afetivo-emocional (Hotopf *et al.*, 1998; Hotopf *et al.*, 1999; Jones *et al.*, 2007).

Em ambiente hospitalar, queixas de dor que ultrapassam o tempo de cura esperado podem estar presentes em até 10% das crianças hospitalizadas (Huguet e Miro, 2008; Taylor *et al.*, 2008; Stevens *et al.*, 2012; Zhu *et al.*, 2012; Friedrichsdorf *et al.*, 2015). No entanto, apesar de sua prevalência (King *et al.*, 2011), uma proporção significativa de crianças hospitalizadas tem a sua dor subavaliada e subtratada (Helgadottir, 2000; Taylor *et al.*, 2008; Johnston *et al.*, 2011; Stevens *et al.*, 2011; Stevens *et al.*, 2012; Friedrichsdorf *et al.*, 2015).

Para o tratamento eficaz das diversas condições de dor crônica é imprescindível que a equipe de saúde disponha de ferramentas de avaliação precisas, confiáveis e de fácil aplicação. A carência ou imprecisão de critérios ou instrumentos de avaliação dificulta a identificação, monitoramento e manejo correto da dor. Esta dificuldade se torna ainda mais crítica em condições de incapacidade de comunicação, como por exemplo, nas desordens do Sistema Nervoso Central (Hauer e Houtrow, 2017; Warlow e Hain, 2018).

Algumas ferramentas estão disponíveis para a avaliação da dor crônica na pediatria (Saul *et al.*, 2016a; Lee *et al.*, 2017; B. Collins, 2019; Birnie *et al.*, 2019). Porém, é necessário que o instrumento seja confiável e validado para avaliação do construto, dentro da realidade na qual se pretende aplicar (Snyder *et al.*, 2007). Apesar da variedade de instrumentos, a maioria se concentra apenas em uma característica da dor, não havendo uma avaliação completa com informações críticas que permitam o conhecimento da experiência dolorosa, a causa mais provável e o tratamento apropriado (Manworren e Stinson, 2016). Existe uma grande quantidade de literatura

que aborda sobre avaliação e tratamento da dor crônica (Huguet *et al.*, 2010; Manworren e Stinson, 2016; Lee *et al.*, 2017; B. Collins, 2019; Birnie *et al.*, 2019), mas poucos estudos se referem à crianças hospitalizadas (Baldrige *et al.*, 2018).

Além disso, as informações que existem não são efetivamente transferidas para a prática durante a internação (Howard, 2003), devido a deficiência de conhecimento durante a formação profissional, ausência de capacitações, além da alocação limitada e deficiente de recursos para a avaliação da dor, principalmente nos países em desenvolvimento (Fazeny *et al.*, 2000; Gilson *et al.*, 2007; Albertyn *et al.*, 2009; Sanya *et al.*, 2014).

Outra problemática é a carência de estudos que auxiliem na utilização de instrumentos de avaliação de alta confiabilidade, permitindo o manejo da dor de forma correta, segura e baseada em evidências (Howard, 2003). Neste sentido, são necessários estudos que avaliem as propriedades de medida destes instrumentos, antes de sua aplicação, a fim de garantir a identificação, quantificação e qualificação correta da dor crônica, caso esteja presente (Frost *et al.*, 2007; Salmond, 2008; Morkink *et al.*, 2010). A presente revisão sistemática tem como objetivo sintetizar e analisar a qualidade das propriedades de medida de instrumentos aplicados na avaliação de dor crônica em crianças hospitalizadas.

2. REFERENCIAL TEÓRICO

2.1 Dor em crianças

As crianças são capazes de perceber a dor desde o estágio inicial de sua vida pré-natal até o desenvolvimento do sistema nervoso, sendo percebida como um desconforto grave e persistente (Ebrahimpour *et al.*, 2019). Pacientes que vivenciam a dor podem apresentar transtorno de humor, ansiedade, tristeza, raiva e isolamento social, afetando diretamente o seu bem-estar e suas atividades diárias, principalmente quando são crianças, onde há maior dificuldade para entender suas condições de dor (Sallfors *et al.*, 2002; Davern *et al.*, 2007; Forgeron *et al.*, 2010; Lewandowski *et al.*, 2010).

As crianças costumam caracterizar a dor através das fontes causais. Assim, podem conhecer estas fontes como internas, externas e emocionais. As fontes externas são mais facilmente identificáveis, uma vez que podem ver e justificar a causa. Entre elas estão: quedas, cortes e queimaduras (Ebrahimpour *et al.*, 2019). Muitas vezes as fontes emocionais e internas são de difícil entendimento pelas crianças. Elas conseguem associar a dor à doença, mas não entendem como a doença provoca a dor (Kashikar-Zuck *et al.*, 2005).

O processo doloroso é visto pelas crianças como preocupante pois, principalmente em idade pré-escolar, sentem medo de ser feridas e percebem a dor como um sinal de que há algo errado com o seu corpo, podendo ou não ser uma lesão. Outro fator que também influencia é a associação à movimentos e atividades restritas, que os impede de jogar ou ir à escola, sendo determinada como limitante (Perquin *et al.*, 2003; Degotardi *et al.*, 2006). Também ocorre a interpretação como possível punição ou tortura, temendo que tenham cometido erros ou pecados (Jongudomkarn *et al.*, 2006).

Durante a infância, ocorrem mudanças significativas no que se refere às habilidades cognitivas e metacognitivas, necessárias para que a criança compreenda a dor e forneça informações confiáveis (Chan e Von Baeyer, 2016). Quando estas informações não são repassadas ou não são compreendidas, a dor vivenciada na infância pode se cronificar, sendo persistente e causando sintomas físicos e psicológicos até a idade adulta (Walker *et al.*, 2010; Palermo *et al.*, 2019).

A incapacidade de realizar um manejo adequado da dor na criança pode gerar consequências significativas em relação à saúde futura da população e à carga da assistência social à saúde (Connolly *et al.*, 2019). O conhecimento dos fatores relacionados à dor pediátrica permite identificar os maiores riscos que levam ao seu desenvolvimento, o que possibilita uma avaliação e intervenção precoce, e significativo impacto nas taxas de prevalência (King *et al.*, 2011).

2.2 Dor crônica em crianças

A dor crônica é definida por uma dor que dura por três meses ou mais, podendo ser contínua ou recorrente (Huguet e Miro, 2008; King *et al.*, 2011). Pode ser caracterizada por alterações do tecido, alterações do sistema nervoso, patologias pré-existentes, distúrbios de dor primária e síndrome da dor regional complexa (Schechter, 2014; Friedrichsdorf *et al.*, 2016). É considerada um fenômeno multidimensional e nenhum aspecto por si só pode defini-la de forma completa (Ebrahimpour *et al.*, 2019).

A dor crônica em pediatria causa redução da qualidade de vida das crianças (Palermo *et al.*, 2018). Respostas emocionais como ansiedade (Goffaux *et al.*, 2011), depressão, medo-evitação e catastrofização (visualização da dor como pior resultado possível) (Roth *et al.*, 2011), são comuns em pacientes que podem ter cronificação da dor ou já sofrem de dor crônica (Pergolizzi *et al.*, 2014). Além das consequências emocionais, as crianças que vivenciam a dor crônica também necessitam de serviços de saúde com maior frequência, como medicina complementar e alternativa, além dos serviços de urgência, fazendo maior utilização de terapias medicamentosas e aumentando os custos na saúde pública (Tumin *et al.*, 2018).

O impacto da dor crônica também afeta diretamente os pais, uma vez que gera queda na produtividade, além de altos custos para os familiares (Palermo *et al.*, 2018). O tempo gasto para identificação do diagnóstico e tratamento é crucial para a forma que a criança e os familiares vivenciam e compreendem a dor. A experiência incontrolável, assustadora e adversa gera medo e frustração, que aumenta com diagnósticos imprecisos, devido à dificuldade de interpretar a dor durante a avaliação pediátrica (Eccleston e Malleson, 2003; Konijnenberg *et al.*, 2005).

A Organização Mundial de Saúde tem apresentado uma atenção especial à experiência de dor crônica vivenciada pela criança, uma vez que ela pode ser

negligenciada devido à possível incapacidade de se comunicar ou dificuldades de expressar a sua dor de forma clara (WHO Guidelines Approved by the Guidelines Review Committee, 2012). O desafio é saber como as crianças podem se comunicar de forma efetiva, com seus cuidadores e profissionais, para relatar a sua experiência dolorosa, de forma que sejam adequadamente entendidas e cuidadas (Ebrahimpour *et al.*, 2019).

2.3 Dor crônica e hospitalização

A dor pode ser descrita pelas crianças, de acordo com as suas experiências. Quando são saudáveis e com vivências limitadas de dor, consideram os aspectos físicos, como causa. Já crianças que passaram por experiências de dor incomuns, como em situações de hospitalização, podem descrever também aspectos emocionais e psicológicos (Kortessluoma, Nikkonen, *et al.*, 2008; Kortessluoma, Punamaki, *et al.*, 2008).

Dados epidemiológicos têm demonstrado uma prevalência de 16 crianças com queixas ou apresentação clínica de dor crônica a cada 1000 hospitalizadas (Coffelt *et al.*, 2013), e esta realidade vem crescendo a cada ano. Nos Estados Unidos tem se observado um aumento no número de crianças hospitalizadas com diagnóstico de dor crônica, entre os anos de 2004 e 2010. Já no Reino Unido, 22% dos médicos relataram um aumento da incidência desta condição. Assim, pode-se notar que a cronificação da dor tem se tornado um achado bastante frequente durante a internação hospitalar (Bhatia *et al.*, 2008; Coffelt *et al.*, 2013).

Pacientes que se encontram hospitalizados e com dor crônica passam, em média, por 3,2 procedimentos invasivos para um possível diagnóstico e tratamento (Coffelt *et al.*, 2013). A presença de dor moderada a grave é comum em procedimentos e doenças de base, porém diversas vezes, é desconsiderada pela equipe multiprofissional (Hurley-Wallace *et al.*, 2019), devido a limitação de conhecimento e ausência de protocolos institucionais (Lioffi *et al.*, 2018). Quando a criança vivencia a dor, durante a hospitalização, passando por tratamentos, intervenções e ambientes distintos, o seu conceito pode ser influenciado por estes eventos estressantes. Desta forma, ocorre a interpretação no sentido da hospitalização e não da dor especificamente (Ebrahimpour *et al.*, 2019).

Diagnósticos comuns entre crianças hospitalizadas que se queixam de dores crônicas são ansiedade, depressão, comorbidades gastrointestinais e de natureza musculoesquelética (Coffelt *et al.*, 2013). Muitas vezes o manejo da dor crônica no ambiente hospitalar auxilia na redução da frequência e duração de episódios, porém a preocupação durante a assistência ainda permanece voltada para cuidados intensivos, onde o paciente é visto apenas como doença e sintoma, fazendo com que o tratamento não tenha efetividade a longo prazo (Lanzi *et al.*, 2007; Coffelt *et al.*, 2013).

Tem-se observado um aumento na demanda por assistência qualificada em dor crônica, que não tem sido acompanhando pela capacitação profissional (Beckett *et al.*, 2015). Embora exista um avanço nas pesquisas voltadas para o manejo da dor crônica pediátrica nos serviços de saúde, estes conhecimentos não são aplicados na prática assistencial (Kavanagh e Watt-Watson, 2007). Uma variedade de sintomas e queixas associada com a falta de padrões na prática clínica, leva a deficiência, no que se refere à abordagem diagnóstica e manejo da dor crônica infantil (Konijnenberg *et al.*, 2004; Kashikar-Zuck, 2006).

Além disso, existe uma limitação nos recursos validados e qualificados que estejam padronizados e disponíveis para a utilização dos profissionais no manejo da dor em crianças hospitalizadas (Liossi *et al.*, 2019). Desta forma, existe a necessidade de uma estrutura hospitalar adequada com protocolos, associada ao conhecimento multiprofissional para que todos possam compreender e aplicar uma abordagem biopsicossocial da dor (Hurley-Wallace *et al.*, 2019).

2.4 Instrumentos de avaliação da dor

No manejo da dor pediátrica, diversas dimensões podem ser avaliadas, como: sensorial (intensidade, frequência), afetivo/cognitivo (emocional) e impacto da dor no cotidiano (funcionalidade). Embora seja importante conhecer todos os domínios, o parâmetro mais utilizado na prática clínica, visando diagnóstico e plano terapêutico, é a medida de intensidade (Eccleston *et al.*, 2002).

Entre os instrumentos utilizados, o autorrelato é considerado o melhor método para avaliação da dor em pediatria. Porém, muitas vezes, os instrumentos são ofertados para as crianças de forma independente, sem referência sobre a

correspondência entre o grau de demanda envolvida e a capacidade cognitiva da criança no momento da avaliação (Chan e Von Baeyer, 2016; Lee *et al.*, 2017). Desta forma uma avaliação através do autorrelato confiável pode representar um desafio, uma vez que o desenvolvimento cognitivo sofre mudanças durante a infância (Birnie *et al.*, 2019).

Em casos específicos como déficit cognitivo e baixa faixa etária, a dor pode ser avaliada pelo relato do observador, por alterações comportamentais ou fisiológicas (Von Baeyer e Spagrud, 2007), porém outros observadores como pais, cuidadores e profissionais podem ser influenciados durante a avaliação por suas próprias crenças e fatores contextuais (Craig *et al.*, 2010). Desta forma é sugerido que o relatório dos pais entre como informações complementares, não sendo recomendado como substituto do autorrelato, com exceção de casos em que a criança não esteja com a sua capacidade cognitiva adequada para realizar a avaliação (Palermo *et al.*, 2018).

A Escala Numérica de Avaliação -11 (NRS-11), Escala Visual Analógica (EVA), Escala de Dor Facial – Revisada (FPS-R), Escala Analógica de Cor (CAS), Escala de dor Wong-Baker FACES e a Escala Face, Pernas, Atividade, Choro e Consolabilidade (FLACC) são alguns dos instrumentos mais utilizados para avaliação de dor crônica na pediatria (Castarlenas *et al.*, 2017; Lee *et al.*, 2017; Birnie *et al.*, 2019). A NRS-11 pode ser utilizada para avaliar a dor sentida na última semana, através de seus escores que variam de 0 (“sem dor”) a 10 (“pior dor imaginável”) (Chiarotto *et al.*, 2018). A EVA é composta por uma linha horizontal de 100mm de comprimento, com a pontuação que varia de 0 a 10, onde a criança marca um ponto na linha representando a intensidade da sua dor (Huguet *et al.*, 2010; Birnie *et al.*, 2019).

A FPS-R é composta por seis faces, com âncoras que variam de “sem dor” a “muita dor” e cada face possui uma numeração correspondente, variando de 0 a 10, em números pares. As crianças são solicitadas a indicar qual face representa a dor vivenciada por elas (Hicks *et al.*, 2001; Huguet *et al.*, 2010). A CAS utiliza graduações de cor, área e comprimento para que a criança avalie as variações na intensidade da dor vivenciada. Suas graduações vão de branco (“sem dor”), até uma condição mais larga e vermelha (“mais dor”). A criança move o controle que desliza ao comprimento da escala, indicando o seu grau de dor. Por trás do instrumento existe uma escala numérica correspondente de 0 a 10, que indica o número que condiz com a avaliação da criança (Mcgrath *et al.*, 1996; Birnie *et al.*, 2019).

A Escala de dor Wong-Baker FACES é composta por seis faces, que variam de um rosto sorridente ao rosto choroso, e cada face possui uma numeração referente. A classificação numérica varia de 0 (sem machucado) a 10 (pior machucado). Neste instrumento, as crianças são solicitadas para escolher o rosto que mais caracteriza a sua dor (Wong e Baker, 1988). A FLACC é baseada em comportamentos observacionais de dor, onde cada uma das suas cinco categorias (Face, Pernas, Atividade, Choro e Consolabilidade) pode ser pontuada de 0 a 2, resultando em uma pontuação total de 0 a 10 (Hicks *et al.*, 2001; Willis *et al.*, 2003).

2.5 Propriedades de medida

Devido ao aumento do número de instrumentos para avaliar um construto em particular, existe a necessidade de avaliar suas propriedades de medida e conhecer qual o mais adequado para a necessidade da pesquisa ou prática clínica. Quando ocorre a deficiência na qualidade do instrumento, existe a possibilidade dos resultados serem imprecisos e tendenciosos, levando a decisões inapropriadas (Mokkink *et al.*, 2010; Terwee *et al.*, 2012).

As propriedades de medida estão agrupadas em três domínios, os quais são considerados relevantes em instrumentos de avaliação. São eles: confiabilidade, validade e responsividade (Mokkink *et al.*, 2016). A confiabilidade é caracterizada pela capacidade do instrumento de reproduzir um resultado de maneira consistente, no tempo e espaço, ou a partir de diferentes observadores (Terwee *et al.*, 2007). Pode ser afetada por diversos fatores como a função do instrumento, os avaliadores, a população em que é aplicada, o contexto e as circunstâncias. Desta forma a confiabilidade de uma escala não é fixa, e o mesmo instrumento não pode ser considerado confiável em situações diversas (Keszei *et al.*, 2010). Esta propriedade pode ser medida através da consistência interna, confiabilidade e erro de medida (Mokkink *et al.*, 2010).

A consistência interna pode ser definida como o grau em que os itens de um instrumento medem o mesmo construto, levando em consideração a inter-relação entre estes itens (Cortina, 1993; Terwee *et al.*, 2007; Mokkink *et al.*, 2010). Geralmente, é avaliada através do Coeficiente alfa (α) de *Cronbach*, e uma alta

correlação significa que os itens medem o mesmo construto, satisfazendo a avaliação de confiabilidade (Keszei *et al.*, 2010; Streiner e Kottner, 2014).

A confiabilidade também é medida através de avaliação teste-reteste (um instrumento aplicado no mesmo grupo em momentos diferentes), intra-avaliadores (um instrumento aplicado pelo mesmo avaliador em momentos diferentes) ou inter-avaliadores (um instrumento aplicado por diferentes avaliadores no mesmo momento), sendo avaliada através do Coeficiente de Correlação Intraclasse (ICC) (Heale e Twycross, 2015). Já o erro de medida pode ser caracterizado pelo erro sistemático ou aleatório da pontuação de um paciente que não condiz com alterações verdadeiras no construto que está sendo medido (Mokkink *et al.*, 2010). O índice mais utilizado para avaliar esta propriedade é o erro padrão (Mokkink *et al.*, 2010).

A validade de um instrumento se refere a capacidade de mensuração de um construto, podendo ser classificada em validade de conteúdo, validade de critério e validade de construto (Mokkink *et al.*, 2010; Polit, 2015). A validade de conteúdo se refere ao grau em que o conteúdo reflete de forma adequada o construto a ser medido (Polit, 2015). É composta pela validade de face, onde a aparência do instrumento, dificuldade de interpretação dos itens e adequação desejada é avaliada (Zamanzadeh *et al.*, 2015). Geralmente, a avaliação desta propriedade de medida ocorre através de um comitê de especialistas e, posteriormente, através do Índice de Validade de Conteúdo (IVC) é medida a proporção de concordância entre os avaliadores sobre os aspectos do instrumento (Kimberlin e Winterstein, 2008). O índice de concordância aceitável entre os avaliadores especialistas deve ser, no mínimo, 0,80 (Polit e Beck, 2006).

A validade de critério pode ser definida como o grau em que os itens de um instrumento são capazes de refletir adequadamente um outro instrumento que tem a mesma finalidade de medição, caracterizado como “padrão ouro” (Mokkink *et al.*, 2010). Pode ser preditiva (aplicação dos dois instrumentos comparados com um intervalo de tempo) ou concorrente (aplicação dos dois instrumentos concomitantemente) (Kimberlin e Winterstein, 2008; Mokkink *et al.*, 2010). Esta validade pode ser verificada através do coeficiente de correlação, onde as pontuações de medidas dos dois instrumentos são comparadas e quanto mais próximo de 1,00 o coeficiente se encontra, maior a correlação existente (Keszei *et al.*, 2010).

A validade de construto pode ser caracterizada pelo grau em que os escores do instrumento apresentam consistência, com base na construção de hipóteses, se baseando no pressuposto de que o instrumento é eficaz na medida de um determinado construto. Pode ser classificada em validade transcultural, teste de hipótese e validade estrutural (Mokkink *et al.*, 2010).

A validade transcultural pode ser caracterizada pela capacidade dos itens, após uma tradução ou adaptação cultural, de ser uma reflexão adequada da sua versão original (Mokkink *et al.*, 2010). Para que esta avaliação seja feita de forma qualificada deve-se seguir os Padrões baseados em Consensus para a seleção de Instrumentos de Medição de saúde (*COnsensus-based Standards for the selection of health Measurement INstruments - COSMIN*), onde deve haver uma tradução e retradução por tradutores independentes, tradução revisada por um comitê de especialistas e pré-teste do instrumento (Mokkink *et al.*, 2016).

O teste de hipótese é definido pelo grau em que as pontuações de um instrumento, através de hipóteses, medem o construto que se pretende (Mokkink *et al.*, 2010). Pode ser realizado pela técnica de grupos conhecidos (Kimberlin e Winterstein, 2008), através da validade convergente, pela correlação entre pontuações de instrumentos que medem o mesmo construto e também pela validade discriminante (Polit, 2015).

A validade estrutural pode ser caracterizada como o grau em que os escores de um determinado instrumento refletem de forma adequada a dimensão do construto a ser medido. É avaliada através da análise fatorial, sendo recomendada que se utilize a Análise Fatorial Confirmatória (CFA). Outra propriedade que é fundamental para saber se o instrumento possui a capacidade de identificar mudanças ao longo do tempo é a responsividade que pode ser avaliada por meio do teste t e do tamanho do efeito (Mokkink *et al.*, 2010).

3. OBJETIVOS

3.1 Objetivo geral

Sintetizar e analisar a qualidade das propriedades de medida de instrumentos aplicados na avaliação de dor crônica em crianças hospitalizadas.

3.2 Objetivos específicos

- Sintetizar os estudos que abordam sobre propriedades de medida de instrumentos de avaliação da dor crônica em crianças hospitalizadas;
- Avaliar a qualidade metodológica dos estudos através das propriedades de medida dos instrumentos utilizados.

4. METODOLOGIA

A revisão atual foi previamente registrada no Registro Internacional Prospectivo de Revisões Sistemáticas (PROSPERO) sob n. CRD42018099021 e foi conduzida de acordo com as recomendações de Itens de Relatórios Preferidos para Revisões Sistemáticas e Meta-Análises (*Preferred Reporting Items for Systematic reviews and Meta-Analyses* - PRISMA) (Moher *et al.*, 2009). O Protocolo da revisão sistemática (Apêndice I) foi submetido no periódico científico *Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports* (JBISIR).

4.1 Critérios de inclusão

4.1.1 Participantes

Foram incluídos na revisão estudos em que os participantes fossem crianças com idades ≥ 2 anos e < 12 anos, hospitalizadas, com diagnóstico médico de dor crônica, independentemente de sua origem ou queixa de dor persistente (em curso) ou recorrente (episódica) com tempo de duração superior a três meses (Huguet e Miro, 2008; King *et al.*, 2011).

4.1.2 Construto

O construto de interesse para esta revisão é dor crônica em crianças hospitalizadas.

4.1.3 Instrumentos

Foram incluídas ferramentas de avaliação, como questionários, escalas, inventários ou índices utilizados para investigar condições de dor crônica em crianças hospitalizadas. Sendo consideradas todas as formas de avaliação (autorrelato, observação comportamental e alterações fisiológicas). Foram excluídos instrumentos que não avaliam a intensidade da dor.

4.1.4 Resultados

Estudos que incluem, pelo menos, uma propriedade de medida como resultados. As seguintes propriedades de medida foram investigadas:

- Confiabilidade: consistência interna, erro de medida, confiabilidade teste-reteste, confiabilidade inter-avaliador e intra-avaliador.

- Validade: validade de conteúdo, validade de face, validade de construto, validade estrutural, validade transcultural, teste de hipóteses, validade de critério.
- Responsividade: sensibilidade à mudança.

Cada propriedade de medição foi classificada de acordo com os critérios para boas propriedades de medição (Prinsen *et al.*, 2018).

4.1.5 Tipos de estudos

Esta revisão considerou qualquer estudo que tivesse como objetivo projetar e validar um instrumento, ou qualquer estudo que pretendesse avaliar as propriedades de medida de uma ferramenta, previamente desenhada, para avaliar a dor crônica em crianças hospitalizadas.

4.2 Estratégia de busca

A estratégia de busca teve como objetivo identificar todos os estudos, sem limite cronológico e sem limitação de linguagem. Títulos e resumos foram revisados considerando palavras de texto livre e termos de índice dos elementos principais: construto (dor crônica); população (crianças hospitalizadas); cenário (hospitalização); tipo de instrumentos (medição de dor) e propriedades de medição (validade, confiabilidade e responsividade). As bases de dados pesquisadas incluíram: PubMed, EMBASE, CINAHL, Web of Science, PsycNET, Banco de Dados Cochrane para Revisões Sistemáticas e Google Scholar. Os termos de pesquisa para propriedades de medição, disponíveis para Pubmed e EMBASE, foram executados usando o filtro de pesquisa sensível validado de acordo com o site COSMIN (Cosmin, 2019).

Listas de referências de todos os artigos incluídos na revisão também foram pesquisadas para estudos adicionais. Não foi considerado banco de dados para literatura cinza. A estratégia de pesquisa completa para o banco de dados PubMed está detalhada no Apêndice II.

4.3 Seleção de estudos

Após a busca dos estudos todas as citações identificadas foram agrupadas e carregadas no programa *EndNote* versão X6 (Clarivate Analytics, PA, EUA). Os estudos encontrados foram excluídos por duplicação, leitura de títulos e resumos, além da leitura dos artigos na íntegra, seguindo os critérios de inclusão e exclusão. O registro da quantidade de artigos excluídos em cada fase está apresentado em um fluxograma nos resultados.

4.4 Avaliação da qualidade metodológica

A qualidade dos artigos selecionados foi avaliada, pontuada e qualificada usando o *checklist* Risco de Viés de acordo com as recomendações do COSMIN (Anexo I) (Mokkink *et al.*, 2018). O *checklist* COSMIN é uma ferramenta padronizada para avaliar a qualidade metodológica dos estudos com base nas propriedades de medida, como consistência interna, confiabilidade, erro de medida, validade de conteúdo, validade estrutural, teste de hipóteses, validade transcultural, validade de critério e responsividade. A qualidade de cada estudo pode ser classificada em muito boa, adequada, duvidosa ou inadequada. Para determinar esta classificação, foi levado em consideração a menor classificação, dentre os itens de cada propriedade de medida avaliada (Mokkink *et al.*, 2018).

Cada propriedade também foi classificada de acordo com os critérios para boas propriedades de medição (Anexo II). As propriedades podem ser classificadas como suficiente (+), insuficiente (-) ou indeterminada (?). Posteriormente, os resultados foram sumarizados e avaliados, também, de acordo com os critérios de qualidade para boas propriedades de medição, podendo ser suficiente (+), insuficiente (-), inconsistente (\pm) ou indeterminado (?) (Terwee *et al.*, 2007). Por fim, a qualidade da evidência foi avaliada utilizando a abordagem GRADE do COSMIN (Anexo III), que especifica quatro níveis de classificação: alta, moderada, baixa e muito baixa. A abordagem GRADE pode rebaixar a qualidade das evidências quando existe uma preocupação sobre a confiabilidade dos resultados, levando em consideração quatro fatores: risco de viés, inconsistência, imprecisão e indirectividade (Prinsen *et al.*, 2018).

O ponto de partida é sempre supor que a qualidade das evidências tem uma classificação alta e que pode ser rebaixada em um nível (sério) ou dois (muito sério), em cada fator, podendo ser até três níveis (extremamente sério risco de viés) quando a evidência é baseada em apenas um estudo inadequado (Prinsen *et al.*, 2018). A avaliação da qualidade metodológica foi realizada por dois pesquisadores independentes e um terceiro pesquisador, quando necessário, arbitrou as diferenças, conforme recomendações do COSMIN (Mokkink *et al.*, 2018).

4.5 Extração e síntese de dados

Os dados foram extraídos de artigos incluídos na revisão usando uma ferramenta de extração de dados padronizada do Sistema para o Gerenciamento Unificado, Avaliação e Revisão de Informações do Instituto Joanna Briggs (JBI SUMARI) (Munn *et al.*, 2019), adaptada aos elementos específicos de uma revisão psicométrica (Apêndice III).

Os dados extraídos incluíram: características do estudo (título, ano de publicação, autor, desenho metodológico, objetivos e cenário), características da amostra de inclusão (idade, gênero, características clínicas e tamanho da amostra), características das ferramentas de avaliação da dor (nome, modo de administração, avaliador, respondente, capacidade do avaliador e avaliação da capacidade cognitiva da criança em caso de autorrelato), propriedades de medição (validade de conteúdo, validade estrutural, consistência interna, validade transcultural, confiabilidade, erro de medida, validade de critério, validade de construto, responsividade e ferramentas aplicadas para auxiliar no processo de validação) e descrição dos resultados.

A extração dos dados foi realizada por dois pesquisadores independentes e um terceiro pesquisador, quando necessário, arbitrou as diferenças. Se os dados estivessem faltando ou não estivessem claros, os autores entrariam em contato com os pesquisadores para esclarecimentos. Posteriormente, foi realizada uma análise crítica dos dados, inicialmente, de forma independente e depois através de uma reunião de consenso. A análise detalhada está apresentada na seção de discussão, na forma narrativa, além de exposta em tabela para facilitar a interpretação das informações.

5. RESULTADOS

Inicialmente foram encontrados 1.298 artigos. Ao final somente duas publicações preencheram os critérios pré-estabelecidos e foram incluídas na análise. O fluxograma do processo de seleção dos estudos está descrito na Figura 1.

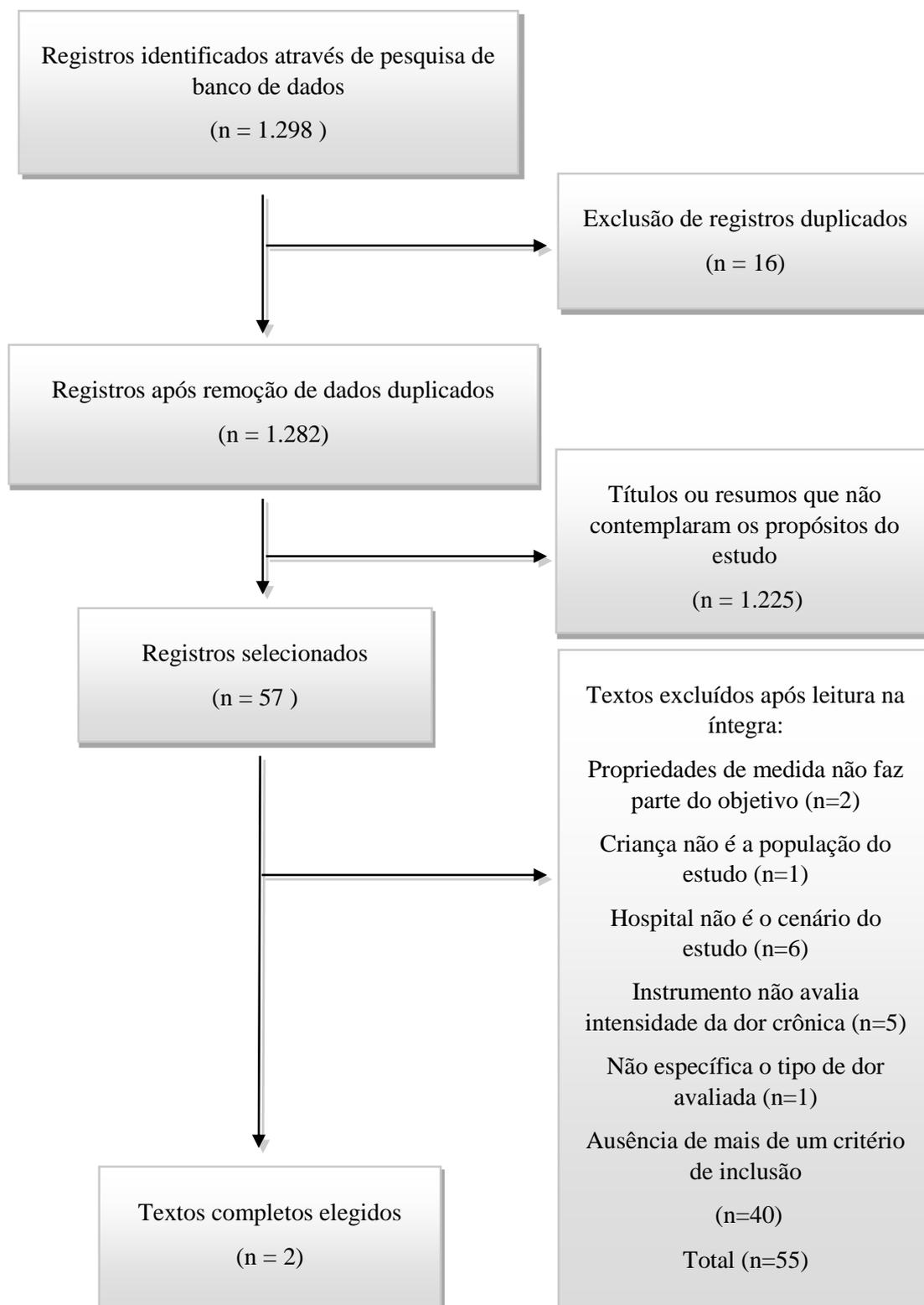


Figura 1. Fluxograma do estudo.

Entre os 55 artigos que foram excluídos após leitura na íntegra, foram encontrados diversos instrumentos que avaliam a intensidade da dor, além de diversos fatores que envolvem o processo doloroso. Conforme mostra a Tabela 1.

Tabela 1. Instrumentos de avaliação da dor encontrados em estudos excluídos após leitura do artigo na íntegra

Instrumentos de avaliação	Estudos
Escala de Comportamento de Conforto (COMFORT)	(Bosch-Alcaraz <i>et al.</i> , 2018)
Escala Analógica de Cor (CAS)	(Bulloch <i>et al.</i> , 2009)
Escala de Autoeficácia Infantil	(Bursch <i>et al.</i> , 2006)
Inventário de Qualidade de Vida em Pediatria (PedsQL)	(Chen <i>et al.</i> , 2007)
Inventário de Deficiência Funcional (FDI)	(Claar e Walker, 2006) (Kashikar-Zuck <i>et al.</i> , 2011) (Offenbacher <i>et al.</i> , 2016) (Stahlschmidt <i>et al.</i> , 2018)
Escala de Catastrofização da Dor (PCS-C)	(Crombez <i>et al.</i> , 2003) (Pielech <i>et al.</i> , 2014)
Inventário Central de Sensibilização (CSI)	(Cuesta-Vargas <i>et al.</i> , 2018)
Iniciativa do Sistema de Informação de Medição de Resultados Relatados pelo Paciente PROMIS	(Cunningham <i>et al.</i> , 2017) (Kashikar-Zuck <i>et al.</i> , 2016) (Jacobson Jr <i>et al.</i> , 2015)
Escala Facial de Dor	(Da Silva <i>et al.</i> , 2011)
Escala Face, Pernas, Atividade, Choro e Conslabilidade (FLACC)	(Da Silva <i>et al.</i> , 2011)
Questionário Medo de Dor (FOPQ-C)	(Dekker <i>et al.</i> , 2018) (Simons <i>et al.</i> , 2011)
Questionário de Estágios de Mudança da Dor (PSOCQ)	(Evans <i>et al.</i> , 2015) (Guite <i>et al.</i> , 2011)

Continuação Tabela 1.

Instrumentos de avaliação	Estudos
Questionário de Pescoço Bournemouth (NBQ-I)	(Geri <i>et al.</i> , 2015)
Entrevista de Limitações das Atividades da Criança	(Hainsworth <i>et al.</i> , 2007) (Palermo <i>et al.</i> , 2004)
Inventário de Enfrentamento da Dor Pediátrica (PPCI)	(Hechler <i>et al.</i> , 2008)
Índice de Dor Abdominal (API)	(Laird <i>et al.</i> , 2015)
Índice de Interferência da Dor (PPI)	(Martin <i>et al.</i> , 2015)
Questionário de Aceitação da Dor Crônica (CPAQ-A)	(Mccracken <i>et al.</i> , 2010)
Questionário de Experiência de Injustiça (IEQ)	(Miller <i>et al.</i> , 2016)
Entrevista de Limitações das Atividades da Criança-21 (CALI-21)	(Palermo <i>et al.</i> , 2008)
Questionário de Enfrentamento da Dor (PCQC)	(Reid <i>et al.</i> , 1998)
Ferramenta de Triagem de Dor Pediátrica (PPST)	(Sil <i>et al.</i> , 2019) (Simons <i>et al.</i> , 2015)
Questionário de Crenças de Dor, abreviado (PBQ-SF)	(Stone <i>et al.</i> , 2016)
Pontuação de Resultados Relacionado a ICF (Classificação Internacional de Funcionamento)	(Van De Voorde <i>et al.</i> , 2011)
Classificação da Dor Crônica (CPG)	(Wager <i>et al.</i> , 2013)
Inventário de Resposta a Dor (PRI)	(Walker <i>et al.</i> , 2008)
Modelo de Avaliação e Enfrentamento da Dor	(Walker <i>et al.</i> , 2005)

5.1 Análise da qualidade metodológica nos estudos através do *checklist* Risco de Viés COSMIN.

A avaliação realizada através da utilização do *checklist* Risco de Viés COSMIN se encontra na Tabela 2. Apesar do desenvolvimento de instrumento não ser considerado uma propriedade de medida o COSMIN avalia esta condição, assim no estudo de Hermann *et al.* (2008), o desenvolvimento do instrumento de avaliação Questionário de Experiência da Dor para Crianças (PEQ-C) apresentou qualidade metodológica classificada como duvidosa, pois os autores não deixam claro se o grupo de moderadores/ entrevistadores foram treinados ou se possuíam experiência. Também não informam se, pelo menos, 10% dos dados foram codificados por dois pesquisadores independentes, não informam se a saturação dos dados foi atingida, além de não estar claro se os pacientes foram questionados sobre a compreensibilidade do questionário. Já a validade estrutural, neste estudo, apresentou qualidade metodológica classificada como muito boa.

No estudo de Marec-Berard *et al.* (2015), a qualidade metodológica do desenvolvimento do instrumento Avaliação Hétero Douleur Enfant (HEDEN) foi classificada como inadequada, uma vez que o estudo não relata sobre a realização de um estudo de entrevista cognitiva ou teste piloto do instrumento desenvolvido. No que se refere às propriedades de medida, o estudo avalia algumas propriedades do instrumento HEDEN em duas etapas diferentes, desta forma foi considerada, na revisão, cada etapa isoladamente, assim a validade estrutural, consistência interna, validade de critério e validade de construto apresentaram qualidade metodológica classificada como muito boa, em todas as etapas avaliadas. Já a confiabilidade apresentou qualidade metodológica classificada como inadequada, uma vez que coeficiente kappa não foi calculado.

5.2 Avaliação de acordo com os critérios para boas propriedades de medição.

O resultado de cada estudo isolado foi avaliado de acordo com os critérios para boas propriedades de medição (Terwee *et al.*, 2007). Cada resultado foi classificado como suficiente (+), insuficiente (-) ou indeterminado (?). Conforme apresenta a Tabela 3. No estudo de Hermann *et al.* (2008), a validade estrutural de PEQ-C apresentou-se como suficiente, uma vez que CFA: Índice de Ajuste Comparativo

(CFI)=0,98 e Erro Quadrático Médio da Raiz da Aproximação (RMSA)=0,026. A consistência interna também apresentou classificação suficiente, pois o Alfa (α) de *Cronbach* de cada subescala foi 0,85 (gravidade da dor), 0,87 (interferência), 0,81 (angústia) e 0,71 (apoio social).

No estudo de Marec-Berard *et al.* (2015), a validade estrutural (2º etapa) apresentou classificação insuficiente, pois a CFA não foi avaliada, apenas a análise fatorial exploratória. No que se refere a consistência interna, a 1º etapa apresentou classificação indeterminada para esta propriedade de medida, uma vez que a validade estrutural não foi avaliada. Na 2º etapa a consistência interna também apresentou classificação indeterminada, pois a validade estrutural apresentou classificação insuficiente.

A confiabilidade apresentou classificação insuficiente, pois na 1º etapa ICC=0,62, e na 2º etapa ICC=0,67. A validade de critério apresentou classificação insuficiente, sendo encontrada uma baixa correlação entre os itens de HEDEN e suas contrapartes na Escala Dor da Criança Gustave Roussy (Gauvain-Piquard *et al.*, 1999), tanto na 1º etapa (0,27- 0,41), quanto na 2º etapa (0,60). Por fim, a validade de construto apresentou classificação indeterminada, nas duas etapas, pois não foi definida nenhuma hipótese pela equipe de revisão (Marec-Berard *et al.*, 2015).

5.3 Avaliação das evidências sumarizadas e classificação conforme abordagem GRADE

Os resultados foram sumarizados por cada propriedade de medição de cada instrumento avaliado nos estudos selecionados em uma Tabela Resumo da Descoberta (*Summary of Finding - SoF*). O resumo dos resultados também foi classificado de acordo com os critérios de qualidade para boas propriedades de medição, podendo ser suficiente (+), insuficiente (-), inconsistente (\pm) ou indeterminado (?) (Terwee *et al.*, 2007). Posteriormente, a classificação foi realizada utilizando a abordagem GRADE (Prinsen *et al.*, 2018). Conforme apresenta a Tabela 4.

Após resumir os resultados encontrados, a classificação geral das propriedades de medida, por instrumento, e segundo os critérios para boas propriedades de medição, foram iguais aos resultados dos estudos isolados, pois segundo Prinsen *et*

al. (2018), quando as classificações, por estudo, são todas suficientes, insuficientes ou indeterminadas, os resultados agrupados também apresentarão a mesma classificação. Como ocorreu com a validade estrutural, que se apresentou suficiente no PEQ-C e insuficiente na HEDEN, a consistência interna, que se apresentou como suficiente no PEQ-C e indeterminado na HEDEN, além da confiabilidade e validade de critério que se apresentaram insuficiente na HEDEN, e validade de construto que foi classificada com indeterminada na HEDEN.

No que se refere à abordagem GRADE, a validade estrutural foi avaliada em dois instrumentos, no PEQ-C a evidência apresentou classificação moderada, com sério risco relacionado à indirectividade, pois no estudo que abordou esta ferramenta apenas parte da população se encontra hospitalizada (Hermann *et al.*, 2008). Na escala HEDEN a classificação da validade estrutural foi muito baixa, com risco muito sério para imprecisão, pois esta propriedade foi avaliada apenas na segunda etapa do estudo, com uma amostra de 48 crianças (Total $n < 50$), e risco sério na indirectividade, pois parte da população é composta por crianças com dor aguda (Marec-Berard *et al.*, 2015).

Outra propriedade de medida avaliada em dois instrumentos foi a consistência interna. No PEQ-C a classificação pela abordagem GRADE foi moderada, pois apresentou risco sério na indirectividade, devido parte da população do estudo se encontrar hospitalizada (Hermann *et al.*, 2008). Na escala HEDEN, o resumo dos estudos apresentou classificação moderada, com risco sério na indirectividade, devido uma parcela da população ser de crianças com dor aguda (Marec-Berard *et al.*, 2015). A confiabilidade foi avaliada apenas na escala HEDEN, onde o resumo dos estudos apresentou classificação muito baixa, segundo a abordagem GRADE, pois apresentou risco de viés muito sério, por haver a presença de dois estudos (2 etapas) com qualidade inadequada, além do risco sério para indirectividade, uma vez que parte da população se apresenta com dor aguda (Marec-Berard *et al.*, 2015).

A validade de critério foi avaliada apenas na escala HEDEN apresentando classificação moderada, segundo a abordagem GRADE, com risco sério para indirectividade, pois uma parcela da população se apresenta com dor aguda (Marec-Berard *et al.*, 2015). A validade de construto também foi avaliada apenas na escala HEDEN onde apresentou classificação baixa, com risco sério na imprecisão, devido amostra total < 100 ($n = 59$), uma vez que foi avaliada apenas na 1º etapa do estudo,

além do risco sério para indirectividade, por uma parcela da população ser composta de crianças com dor aguda (Marec-Berard *et al.*, 2015).

5.4 Dados extraídos dos estudos incluídos na revisão

Conforme a Tabela 5. Os dados dos estudos foram extraídos seguindo as características dos estudos, características da amostra de inclusão, características das ferramentas de avaliação da dor crônica, propriedades de medição e descrição dos resultados.

5.4.1 Características dos estudos incluídos

Os estudos foram publicados nos anos de 2008 (Hermann *et al.*, 2008) e 2015 (Marec-Berard *et al.*, 2015) e apresentam desenvolvimento e validação de instrumento como desenho metodológico. No que se refere aos objetivos os estudos apresentam algumas distinções, onde um artigo pretendeu desenvolver e iniciar a validação de um breve questionário para crianças e pais, utilizados na avaliação da dor crônica em crianças e adolescentes (Hermann *et al.*, 2008), e o segundo estudo teve como objetivo desenvolver e validar uma ferramenta de avaliação da dor mais rápida e fácil de usar à beira do leito do paciente (Marec-Berard *et al.*, 2015). Os dois estudos foram realizados com crianças hospitalizadas, porém, um artigo também abordou crianças durante consultas ambulatoriais (Hermann *et al.*, 2008).

5.4.2 Características da amostra de inclusão

Os dois estudos que foram incluídos na revisão apresentam crianças na sua amostra, sendo uma população de 111 crianças e adolescentes com faixa etária entre 7 e 18 anos (Hermann *et al.*, 2008), e população de 107 crianças (1º etapa com 59 crianças e 2º etapa com 48 crianças) entre 2 e 6 anos (Marec-Berard *et al.*, 2015), sendo crianças de ambos os sexos. Como a existência da dor crônica foi apresentada pelos estudos como critério de inclusão, as crianças apresentaram diversas condições de dor como enxaqueca, cefaleia tensional, fibromialgia, artrite reumatoide, dor abdominal recorrente e doença inflamatória intestinal (Hermann *et al.*, 2008), além de dor abdominal pélvica, dor em cabeça e pescoço, dor em região torácica, dor em membros superiores e inferiores (Marec-Berard *et al.*, 2015).

5.4.3 Características das ferramentas de avaliação da dor crônica

Os instrumentos de avaliação da dor aplicados nos estudos foram o PEQ-C (Hermann *et al.*, 2008) e a HEDEN (Marec-Berard *et al.*, 2015). A ferramenta PEQ-C é uma medida de autorrelato, que foi utilizada pela própria criança, sem necessidade de ajuda de outra pessoa (Hermann *et al.*, 2008), já a HEDEN é uma medida de observação comportamental que foi aplicada por uma enfermeira treinada (Marec-Berard *et al.*, 2015).

5.4.4 Propriedades de medição

No que se refere à avaliação das propriedades de medida, o PEQ-C (Hermann *et al.*, 2008) teve a validade estrutural e consistência interna avaliadas, tendo como instrumentos comparativos uma adaptação do Inventário de Somatização Infantil (CSI) (Walker *et al.*, 2009), Teste de Depressão para Crianças (DTC) (Rossmann, 2005), Teste de Ansiedade Infantil II (CAT-II) (Thurner e Tewes, 2000), Questionário de Pontos Fortes e Dificuldades (SDQ) (Goodman, 2001), Inventário dos Filhos Parentais (PCI) (Krohne *et al.*, 1984), Questionário de Práticas de Parentalidade Materna (MPPQ) (Schneewind, 1999) e o Questionário de Cognições Relacionadas à Dor para Crianças e Adolescentes (PRCQ-C) (Hermann *et al.*, 2007). A escala HEDEN (Marec-Berard *et al.*, 2015) teve a validade estrutural, consistência interna, validade de critério, confiabilidade e validade de construto avaliados, tendo como instrumento comparativo a DEGR (Gauvain-Piquard *et al.*, 1999).

5.4.5 Descrição dos resultados

Inicialmente foi desenvolvido o PEQ-C, originado da parte A do Inventário de Dor Multidimensional (MPI) alemão para adultos (Flor *et al.*, 1990), e composto por quatro subescalas que avaliam a gravidade da dor, interferência relacionada à dor, sofrimento afetivo relacionado à dor e suporte social percebido. Durante a análise fatorial exploratória, todos os itens do PEQ-C apresentaram carga fatorial satisfatória, mostrando boa adequação dos dados (Hermann *et al.*, 2008).

Considerando a associação entre subescalas da PEQ-C e os instrumentos aplicados para auxiliar na validação, foi observado que a gravidade da dor medida pela PEQ-C apresentou correlação positiva com a frequência (0,47) e intensidade (0,42) da dor avaliada pelo CSI adaptado, já com o nível de ansiedade da criança avaliado pelo CAT-II (0,17) ou problemas comportamentais avaliado pelo SDQ (0,14),

não houve correlação significativa. A subescala estresse efetivo da PEQ-C se mostrou correlacionada positivamente com nível de sintomas depressivos avaliada pela DTC (0,46-0,57) e ansiedade de traço avaliada pela CAT-II (0,55). O apoio social relacionado à dor da PEQ-C foi associado à percepção do suporte materno geral avaliado pelo ICP (0,40) e comportamento materno afetivo avaliado pelo MPPQ (0,38) (Hermann *et al.*, 2008).

No que se refere à correlação entre as subescalas da PEQ-C, a interferência relacionada à dor apresentou uma correlação significativamente positiva com a gravidade da dor e sofrimento afetivo. Já a subescala de gravidade da dor apresentou uma correlação positiva, mas não significativa, ao sofrimento afetivo. O suporte social percebido apresentou uma correlação positiva, mas não significativa, com interferência relacionada à dor, porém não apresentou relação com o sofrimento afetivo nem com a gravidade da dor. Com relação à consistência interna, todas as subescalas apresentaram-se satisfatórias (Hermann *et al.*, 2008).

A consistência interna das quatro subescalas de PEQ-C correspondeu de perto aos valores para o MPI adulto original (0,70-0,90), encontrados no estudo de Kerns, Turk e Rudy (1985). Além disso, o PEQ-C também apresentou um padrão de intercorrelação de escala consistente com o MPI original, apresentado pelo mesmo estudo, onde a correlação entre gravidade da dor e sofrimento afetivo foram semelhantes (PEQ-C: 0,35, MPI: 0,34). Em comparação com o MPI original a severidade da dor apresentou uma correlação menos forte com interferência relacionada a dor (MPI:0,58, PEQ-C: 0,37). Já as subescalas de interferência e angústia afetiva foram mais correlacionadas no PEQ-C (0,46) do que no MPI (0,26) (Kerns *et al.*, 1985; Hermann *et al.*, 2008).

Com relação à escala HEDEN, o estudo que avaliou suas propriedades de medida realizou a pesquisa em duas etapas. Na 1^o etapa o instrumento apresentou baixa correlação entre os seus itens e os itens da DEGR (0,27-0,41), porém apresentou maior validade convergente (0,50) ao testar os escores totais da HEDEN com os escores totais da DEGR. Com relação à consistência interna o alfa (α) de *Cronbach* apresentou valor de 0,61. A análise fatorial apresentou que o modelo não é apropriado, além de não fornecer uma categorização adequada dos itens. A confiabilidade também foi avaliada, apresentando um ICC=0,62. Na 2^o etapa o coeficiente alfa (α) de *Cronbach* foi 0,75, indicando boa consistência interna. A

correlação entre as pontuações totais de HEDEN e DEGR foi de 0,60. Na Análise fatorial o escore de Kaiser-Meyer-Olkin foi de 0,73, indicando adequação satisfatória. A avaliação da confiabilidade apresentou ICC=0,67 (Marec-Berard *et al.*, 2015).

Tabela 2. Análise de qualidade metodológica nos estudos através do *checklist* Risco de Viés COSMIN

Propriedade de medida	Referência / Instrumento de avaliação	
	Hermann <i>et al.</i> (2008) / PEQ-C	Marec-Berard <i>et al.</i> (2015) / HEDEN
Desenvolvimento de instrumento	Duvidosa	Inadequada
Validade de conteúdo	-----	-----
Validade estrutural	Muito boa	2º Etapa: Muito boa
Consistência Interna	Muito boa	1º Etapa: Muito boa 2º Etapa: Muito boa
Validade transcultural	-----	-----
Confiabilidade	-----	1º Etapa: Inadequada 2º Etapa: Inadequada
Erro de medida	-----	-----
Validade de critério	-----	1º Etapa: Muito boa 2º Etapa: Muito boa
Teste de hipótese para validade de construto	-----	1º Etapa: Muito boa
Responsividade	-----	-----

PEQ-C= Questionário de Experiência da Dor para crianças; HEDEN= Avaliação Hétero Douleur Enfant.

Tabela 3. Classificação de acordo com os critérios para boas propriedades de medição

Propriedade de medida	Referência / Instrumento de avaliação	
	Hermann <i>et al.</i> (2008) / PEQ-C	Marec-Berard <i>et al.</i> (2015) / HEDEN
Validade de conteúdo	-----	-----
Validade estrutural	+	2º Etapa: -
Consistência Interna	+	1º Etapa: ? 2º Etapa: ?
Validade transcultural	-----	-----
Confiabilidade	-----	1º Etapa: - 2º Etapa: -
Erro de medida	-----	-----
Validade de critério	-----	1º Etapa: - 2º Etapa: -
Teste de hipótese para validade de construto	-----	1º Etapa: ?
Responsividade	-----	-----

PEQ-C= Questionário de Experiência da Dor para crianças; HEDEN= Avaliação Hétero Douleur Enfant. Legenda de classificação dos resultados: suficiente (+), insuficiente (-) e indeterminado (?).

Tabela 4. Tabela Resumo da Descoberta (*Summary of Finding - SoF*) e avaliação GRADE

Propriedade de medida	Resumo ou resultado agrupado	Avaliação geral	Qualidade da evidência
Validade de conteúdo	-----	-----	-----
Validade estrutural			
PEQ-C	CFA: CFI =0,98 e RMSA= 0,026, tamanho total da amostra: 111	+	Moderada
HEDEN	Análise fatorial exploratória avaliada (Escore de Kaiser-Meyer-Olkin= 0,73), tamanho total da amostra: 48	-	Muito baixa
Consistência Interna			
PEQ-C	Alfa (α) de Cronbach de subescalas resumido: 0,71-0,87, tamanho total da amostra: 111	+	Moderada
HEDEN	Alfa (α) de Cronbach de esubescalas resumido: 0,61-0,75, tamanho total da amostra: 107	?	Moderada
Validade transcultural			
Confiabilidade			
HEDEN	Coefficiente de Correlação Intraclasse resumido: 0,36-0,79, tamanho total da amostra: 107	-	Muito Baixa

Erro de medida	-----	-----	-----
Validade de critério			
HEDEN	Resumo da correlação com outro instrumento: 0,13-0,64, tamanho da amostra total: 107	-	Moderada
Teste de hipótese para validade de construto			
HEDEN	A hipótese não foi definida, tamanho da amostra total: 59.	?	Baixa
Responsividade	-----	-----	-----

PEQ-C= Questionário de Experiência da Dor para crianças; HEDEN= Avaliação Hétero Douleur Enfant; CFA= Análise Fatorial Confirmatória; CFI= Índice de Ajuste Comparativo; RMSA= Erro Quadrático Médio da Raiz da Aproximação.

Legenda de classificação dos resultados: suficiente (+), insuficiente (-), inconsistente (±) e indeterminado (?).

Tabela 5. Dados extraídos dos estudos classificados para a revisão sistemática

Características do estudo		
Título	O impacto da dor crônica em crianças e adolescentes: desenvolvimento e validação inicial de uma versão infantil e dos pais do Questionário de Experiência da Dor.	Escala de Dor HEDEN: Uma Escala Comportamental Encurtada para Avaliação de dor Prolongado em Câncer ou Pós-Cirúrgico em crianças de 2 a 6 anos.
Ano de publicação	2008	2015
Autores	Hermann <i>et al.</i>	Marec-Berard <i>et al.</i>
Desenho metodológico	Desenvolvimento e validação de instrumento	Desenvolvimento e validação de instrumento
Objetivos	Desenvolver e iniciar a validação externa de um breve questionário para crianças e pais, utilizados na avaliação da dor crônica em crianças e adolescentes.	Desenvolver e validar uma ferramenta de avaliação da dor mais rápida e fácil de usar à beira do leito do paciente.
Cenário	Ambulatório e hospitalar (reumatologia)	Hospitalar (hematologia, cirurgia geral e oncologia)
Características da amostra de inclusão		
Idade	Faixa etária ente 7 e 18 anos	Faixa etária entre 2 e 6 anos
Gênero	Feminino e masculino	Feminino e masculino
Características clínicas dos participantes	Enxaqueca, cefaleia tensional, fibromialgia, artrite reumatoide, dor abdominal recorrente e doença inflamatória intestinal.	Dor abdominal pélvica, dor em cabeça e pescoço, dor em região torácica, dor em membros superiores e inferiores.
Tamanho da amostra	111 crianças e adolescentes	107 crianças (1º etapa = 59) (2º etapa = 48)
Características das ferramentas de avaliação		
Feramenta que se pretende avaliar	Questionário de Experiência da Dor para crianças (PEQ-C)	Avaliação Hétero Douleur Enfant (HEDEN)

propriedades de medida		
Modo de administração	Autorrelato	Avaliação comportamental
Avaliador	Criança (As crianças preencheram o questionário sem precisar de ajuda para compreender os itens)	Enfermeira
Respondente	Criança	Enfermeira
O avaliador passou por treinamento antes de aplicar o instrumento de avaliação?	Não informado	Sim
Houve avaliação da capacidade cognitiva da criança? (Casos de autorrelato)	Não informado	-----

Propriedades de medição

Propriedades de medida avaliada no estudo	<ul style="list-style-type: none"> • Validade estrutural • Consistência interna 	<ul style="list-style-type: none"> • Validade estrutural • Consistência interna • Confiabilidade • Validade de critério • Validade de construto
Ferramentas aplicadas para auxiliar no processo de validação	<ul style="list-style-type: none"> • Adaptação do Inventário de Somatização Infantil (CSI) • Questionário de Cognições Relacionadas à Dor para Crianças e Adolescentes (PRCQ-C) <ul style="list-style-type: none"> ▪ Teste de Depressão para Crianças (DTC) ▪ Teste de Ansiedade Infantil II (CAT-II) ▪ Questionário de Pontos Fortes e Dificuldades (SDQ) 	<ul style="list-style-type: none"> ▪ Dor da Criança Gustave Roussy (DEGR)

-
- Inventário dos Filhos Parentais (PCI)
 - Questionário de Práticas de Parentalidade Materna (MPPQ)
-

Descrição dos resultados

O PEQ-C se mostrou internamente consistente e apresentou correlação positiva com a maioria dos instrumentos aplicados durante a validação (CSI, DTC, CAT-II, ICP e MPPQ).

A análise de validade produziu um padrão de correlações com medidas de depressão, dor, ansiedade, comportamentos infantis e cognições relacionadas a dor consistentes com dados psicométricos do MPI adulto.

Na primeira etapa, o estudo apresentou correlação aceitável entre DEGR e HEDEN, com boa confiabilidade e concordância entre avaliadores. Na segunda etapa foi observada uma correlação significativa entre DEGR e HEDEN. A confiabilidade foi boa, com excelente concordância entre avaliadores.

6. DISCUSSÃO

O objetivo da presente revisão foi sintetizar informações sobre as propriedades de medida de instrumentos de avaliação da dor crônica em crianças hospitalizadas. Conforme apresentado nos resultados, foram encontrados dois estudos que desenvolveram e avaliaram as propriedades psicométricas nesta população (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015). As propriedades de medida mais investigadas foram a validade estrutural e consistência interna (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015). Apenas um estudo avaliou confiabilidade, validade de critério e validade de construto (Marec-Berard *et al.*, 2015). Em relação à qualidade metodológica, seguindo os critérios do *checklist* Risco de Viés COSMIN, os dois estudos (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015) apresentaram classificação muito boa para a maioria das propriedades avaliadas.

A identificação e síntese de estudos sobre propriedades de medida apresentando qualidade satisfatória, por meio de revisões sistemáticas, é de grande relevância, pois fornece importantes informações sobre sensibilidade e especificidade de determinado instrumento, bem como reduz o risco de desperdício de recursos direcionados à instrumentos clínicos de baixa confiabilidade e efetividade (Ioannidis *et al.*, 2014). Para selecionar um instrumento ideal, deve ser levado em consideração estudos de alta qualidade sobre as propriedades de medida na população-alvo, além de revisões sistemáticas de estudos sobre propriedades de medida com uma metodologia sistemática e transparente (Walton *et al.*, 2015; Mokkink *et al.*, 2018).

A avaliação da qualidade metodológica dos estudos, utilizando checklists apropriados, permite conhecer a confiabilidade dos resultados encontrados, ou seja, o risco de viés existente (Mokkink *et al.*, 2018). Os critérios para boas propriedades de medição também fazem parte da avaliação de qualidade das pesquisas, permitindo o conhecimento da qualidade dos resultados de cada artigo encontrado, além de poder detectar falhas no conhecimento das propriedades e auxiliar em estudos de validação (Terwee *et al.*, 2007). Na presente revisão, a validade estrutural foi classificada como suficiente (Hermann *et al.*, 2008), e insuficiente (Marec-Berard *et al.*, 2015), onde a análise fatorial confirmatória foi avaliada apenas em um estudo (Hermann *et al.*, 2008).

No que se refere à consistência interna, o alfa (α) de *Cronbach* foi avaliado nos dois estudos (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015). Na pesquisa de Hermann *et al.* (2008) o valor encontrado foi $\geq 0,70$, associado com “pelo menos evidência baixa para validade estrutural suficiente”, levando à uma classificação suficiente, já no estudo de Marec-Berard *et al.* (2015) esta mesma propriedade se apresentou indeterminada nas duas etapas, devido a ausência de critérios para “pelo menos evidência baixa para validade estrutural suficiente” (1º e 2º etapa) e valor de alfa (α) de *Cronbach* $< 0,70$ (1º etapa), indicando baixa correlação entre itens de uma escala, o que torna o instrumento injustificado (Terwee *et al.*, 2007).

Uma ferramenta de avaliação internamente consistente deve ser alcançada através de boas definições de construto, bons itens e análise fatorial exploratória, seguida de análise fatorial confirmatória (Terwee *et al.*, 2007). A análise fatorial é relevante, pois determina se os itens formam uma escala geral ou não (Streiner, 1994; Floyd e Widaman, 1995). Em condições que existam hipóteses prévias da dimensão ou modelos pré-existentes, a análise fatorial confirmatória deve ser aplicada, caso não existam hipóteses ou modelos prévios, a análise fatorial exploratória pode ser utilizada (De Vet *et al.*, 2005).

A confiabilidade foi avaliada apenas por um estudo (Marec-Berard *et al.*, 2015) que apresentou uma classificação insuficiente nas duas etapas, devido ao Coeficiente de Correlação Intraclasse com valor inferior a 0,70. O ICC é o parâmetro de confiabilidade mais utilizado para medidas contínuas, sendo bem classificado um valor mínimo de 0,70 para uma amostra de, pelo menos, 50 pessoas (Terwee *et al.*, 2007). O estudo de Marec-Berard *et al.* (2015) apresentou uma amostra > 50 apenas na primeira etapa do estudo, sendo a segunda etapa insatisfatória, uma vez que abordou apenas uma população de 48 crianças. O mesmo estudo também avaliou a validade de critério, levando em consideração o instrumento original. Desta forma, foi observada a correlação entre ele e a ferramenta desenvolvida no mesmo estudo. Para que a classificação seja positiva, a correlação deve ser, pelo menos, 0,70 (Terwee *et al.*, 2007), o que não foi encontrado nas duas etapas do estudo que avaliou esta propriedade de medida. Deste modo, a classificação foi considerada insuficiente.

A validade de construto também foi avaliada por Marec-Berard *et al.* (2015). Hipóteses, correlações esperadas entre medidas, ou diferenças esperadas nas

pontuações entre grupos “conhecidos” podem ser consideradas para avaliar esta propriedade de medida. O ideal é que seja o mais específico possível (Terwee *et al.*, 2007), o que não foi observado, uma vez que a pesquisa não apresentou hipóteses pré-estabelecidas e claras, não sendo também definidas pela revisão sistemática. Quando as hipóteses não são específicas o risco de viés é considerável, pois se pode pensar em alternativas que levaram a uma baixa correlação em vez de entender que o instrumento não é válido (Terwee *et al.*, 2007).

Os estudos incluídos na revisão foram sumarizados e avaliados, conforme os critérios de qualidade para boas propriedades de medição e a abordagem GRADE. Estes dados foram apresentados em uma Tabela Resumo da Descoberta (SoF), que fornece um resumo conciso das principais informações necessárias para que um profissional de saúde possa tomar decisão, além de apresentar informações passíveis de recomendação (Guyatt *et al.*, 2011), como por exemplo, se o instrumento de avaliação da dor crônica é qualificado para utilização. Neste momento o foco passou para o instrumento de avaliação, onde a classificação geral de relevância, abrangência e compreensibilidade foram observados, enquanto nas fases anteriores, o foco estava voltado para cada estudo isolado (Terwee *et al.*, 2018).

As classificações gerais dos estudos encontrados apresentaram os mesmos resultados das avaliações individuais e foram acompanhadas pela classificação da qualidade da evidência, utilizando os fatores GRADE, que indicaram o quanto as classificações gerais são confiáveis (Terwee *et al.*, 2018). O sistema GRADE é muito mais do que um sistema de classificação, pois oferece um processo transparente e estruturado para o desenvolvimento e apresentação de evidências para revisões sistemáticas (Guyatt *et al.*, 2011), baseado no risco de viés, inconsistência, imprecisão e indirectividade (Prinsen *et al.*, 2018).

Na presente revisão, foi observado que a qualidade das evidências para a maioria dos resultados se apresentou moderada e muito baixa, principalmente devido ao risco relacionado à indirectividade, por apenas parte da população se encontrar hospitalizada (Hermann *et al.*, 2008) ou parte da população ter dor aguda (Marec-Berard *et al.*, 2015), além do risco relacionado à imprecisão, devido à amostra inferior a 50 crianças (2º etapa) e existência de dois estudos com qualidade inadequada (Marec-Berard *et al.*, 2015). A existência de poucos estudos que abordam, exclusivamente, dor crônica em crianças hospitalizadas pode ser confirmada por

Marec-Berard et al. (2015), que afirma que a avaliação da dor crônica em crianças hospitalizadas ainda é vista como um desafio, uma vez que sinais evidentes de comportamento podem se habituar ou se dissipar com o passar do tempo dificultando uma avaliação confiável.

Neste contexto, torna-se extremamente importante a avaliação das evidências sobre instrumentos de avaliação de dor crônica em crianças hospitalizadas. A síntese dos resultados pode auxiliar os profissionais de saúde no uso racional e eficaz dos instrumentos, de acordo com as recomendações encontradas (Guyatt *et al.*, 2011). Ao tomar decisões, os profissionais não se baseiam apenas pelas melhores estimativas de vantagens e desvantagens, mas também pela confiança nestas estimativas, assim o reconhecimento das limitações das evidências filtram as recomendações. Uma atenção insuficiente à qualidade das evidências pode levar ao surgimento de recomendações e aplicações de instrumentos inadequados, induzindo à prática de uma assistência ineficaz (Guyatt *et al.*, 2008).

A revisão sistemática é relevante no processo de decisão para um manejo adequado da dor em crianças, uma vez que os estudos presentes podem auxiliar na escolha da ferramenta mais adequada para determinada população, pois possuem como objetivo desenvolver instrumentos de avaliação da dor crônica e avaliar suas propriedades de medida (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015). A avaliação destas ferramentas foi realizada em uma população incluindo crianças em ambiente ambulatorial (Hermann *et al.*, 2008) ou hospitalar (Hermann *et al.*, 2008; Marec-Berard *et al.*, 2015), com faixa etária entre 2 e 6 anos (Marec-Berard *et al.*, 2015), e entre 7 e 18 anos (Hermann *et al.*, 2008).

Entre os instrumentos encontrados está o PEQ-C, uma ferramenta relativamente curta, originada com base na parte A do MPI para adultos alemães (Flor *et al.*, 1990). Apresenta como subescalas a severidade da dor, interferência, sofrimento afetivo e apoio social, sendo o primeiro questionário validado a avaliar o impacto psicossocial da dor crônica em crianças e adolescentes em língua alemã (Hermann *et al.*, 2008).

Já a HEDEN foi originada de uma redução da ferramenta DEGR (Gauvain-Piquard *et al.*, 1999), através de um consenso entre três médicos e uma enfermeira especialista em oncologia pediátrica e envolvida no tratamento da dor. A HEDEN

possui dois itens que avaliam a expressão da dor (postura para aliviar a dor e desconfiança ao ser movido), dois itens que avaliam a inércia psicomotora (falta de interesse pelo ambiente e lentidão e escassez de movimento) e um item que explora a expressão voluntária de dor (expressando a dor) (Marec-Berard *et al.*, 2015).

O PEQ-C é um instrumento de autorrelato e, quando possível, este tipo de relato é considerado padrão-ouro no manejo da dor em crianças (Buttner e Finke, 2000; Von Baeyer *et al.*, 2013). Em algumas situações as crianças não possuem capacidade cognitiva suficiente para a aplicação do autorrelato, como crianças em idade pré-escolar. Além disso, podem negar sua dor ou até mesmo exacerbá-la, confundindo a avaliação. Neste momento, ferramentas que possibilitam uma avaliação observacional, como a HEDEN, podem ser aplicadas (Buttner e Finke, 2000; Von Baeyer *et al.*, 2013).

Apesar da observação comportamental ser uma alternativa, nestas condições, os autores que desenvolveram a HEDEN (Marec-Berard *et al.*, 2015) relatam que a utilização de instrumentos observacionais, exclusivamente, não são muito recomendados em situação de dor crônica, uma vez que os sinais podem se dissipar e não ser claros ao longo do tempo (Von Baeyer e Spagrud, 2007). Exceto em condições de exacerbação da dor ao movimento ou quando alguém faz o paciente recordar a dor (em casos de questionamento) (Tsze *et al.*, 2013).

Os dois estudos apresentaram instrumentos rápidos e práticos para aplicação, o PEQ-C se mostrou de fácil utilização (Hermann *et al.*, 2008), e a HEDEN se mostrou aplicável dentro de cinco minutos (Marec-Berard *et al.*, 2015), ambos sendo considerados viáveis para a utilização na prática clínica. Outro ponto importante é que os avaliadores devem estar familiarizados com o instrumento, tornando a avaliação mais confiável, seja através de treinamentos ou por aplicações repetidas. No entanto, esta importância foi levada em consideração por um único estudo (Marec-Berard *et al.*, 2015). Também foram observados melhores resultados na segunda etapa, durante a aplicação da HEDEN, uma vez que esta familiarização havia ocorrido na primeira etapa, além do relato sobre o treinamento dos avaliadores.

A demanda por avaliação qualificada de dor crônica em crianças hospitalizadas é uma realidade. Atualmente, existem muitos instrumentos de avaliação da dor, porém todos possuem condições específicas para o seu uso. A confiabilidade de uma

ferramenta depende do contexto clínico no qual é aplicada, além da qualidade dos critérios de validade da mesma, e de condições que influenciam na escolha do instrumento adequado (Beltramini *et al.*, 2017). O ideal é a utilização de uma ferramenta sensível, confiável, com boa construção, validada, fácil de entender e com boa aplicabilidade em diversas situações clínicas (Stinson *et al.*, 2006). No entanto, poucas escalas possuem fortes critérios de validade e confiabilidade (Beltramini *et al.*, 2017).

Os resultados encontrados corroboram com esta afirmativa, pois mostram a necessidade da realização de mais estudos com menos riscos, que permitam conclusões com base em maior qualidade de evidência. A presente revisão apresenta algumas limitações: (1) baixo número de estudos encontrados e (2) moderado a baixo nível de evidência, o que limita a recomendação dos atuais instrumentos de avaliação de dor crônica em crianças hospitalizadas.

7. CONSIDERAÇÕES FINAIS

O presente estudo buscou identificar, sintetizar e qualificar todos os instrumentos de avaliação da dor crônica em crianças hospitalizadas. Foi observado que, apesar da existência de uma grande quantidade de instrumentos para avaliação da dor, poucos são voltados, especificamente, para a dor crônica em crianças hospitalizadas, além disso as evidências encontradas se mostraram de qualidade moderada a muito baixa. Neste sentido, urge a necessidade de mais estudos sobre o tema, seguindo as recomendações de validação e avaliação das propriedades psicométricas reconhecidas.

A revisão sistemática destas propriedades é de grande relevância para o conhecimento sobre a existência de instrumentos confiáveis, validados e responsivos para avaliar a dor crônica, no ambiente hospitalar, pois auxilia na escolha do instrumento mais adequado, dentro de um determinado cenário e população. Quando a ferramenta utilizada é eficaz, ocorre uma avaliação confiável, o que facilita o processo de escolha, implementação e avaliação de um tratamento mais efetivo.

Por fim, a existência de diversos instrumentos de avaliação da dor crônica não é útil para a prática clínica se eles não possuírem propriedades de medida confiáveis e validadas, e se não forem acompanhados por pesquisas de alta qualidade. É importante enfatizar que os instrumentos de avaliação da dor crônica em pacientes hospitalizados, provenientes de evidências de alta qualidade, precisam ser divulgados e implantados na prática assistencial, o que não tem sido observado em unidades de internação pediátrica.

REFERÊNCIAS

- AL-ANI, M. Z. et al. Stabilisation splint therapy for temporomandibular pain dysfunction syndrome. **Cochrane Database of Systematic Reviews**, n. 1, 2016. ISSN 1465-1858. Disponível em: < <http://dx.doi.org/10.1002/14651858.CD002778.pub3> >.
- ALBERTYN, R. et al. Challenges associated with paediatric pain management in Sub Saharan Africa. **Int J Surg**, v. 7, n. 2, p. 91-3, Apr 2009. ISSN 1743-9159.
- B. COLLINS, A. **Chronic Pediatric Pain Management: a Review of Multidisciplinary Care and Emerging Topics**. 2019.
- BALDRIDGE, S.; WALLACE, L.; KADAKIA, A. The epidemiology of outpatient pain treatment in pediatrics. **J Pain Res**, v. 11, p. 913-921, 2018. ISSN 1178-7090
- BECKETT, K. et al. **A mixed-method study of pain management practice in a UK children's hospital: Identification of barriers and developing strategies to maintain effective in-patient paediatric pain management**. 2015. n/a-n/a.
- BELTRAMINI, A.; MILOJEVIC, K.; PATERON, D. Pain Assessment in Newborns, Infants, and Children. **Pediatr Ann**, v. 46, n. 10, p. e387-e395, Oct 1 2017. ISSN 0090-4481.
- BERTELSEN, A. K.; BELGRADE, M. Postherpetic Neuralgia, Pharmacological and Non-Pharmacological Treatment Options. **Encyclopedia of Pain**, p. 1853-1857, 2007. ISSN 3540439579.
- BHATIA, A. et al. Chronic pain in children in the UK: a survey of pain clinicians and general practitioners. **Paediatr Anaesth**, v. 18, n. 10, p. 957-66, Oct 2008. ISSN 1155-5645.
- BIRNIE, K. A. et al. Recommendations for selection of self-report pain intensity measures in children and adolescents: A systematic review and quality assessment of measurement properties. **Pain**, v. 160, n. 1, p. 5-18, 2019. ISSN 1872-6623 Disponível em: < <http://dx.doi.org/10.1097/j.pain.0000000000001377> >.
- BOSCH-ALCARAZ, A. et al. Validity of contents of a paediatric critical comfort scale using mixed methodology. **Enfermeria Intensiva**, v. 29, n. 1, p. 21-31, Jan-Mar 2018. ISSN 1130-2399. Disponível em: < <Go to ISI>://WOS:000437382200004 >.
- BULLOCH, B. et al. Reliability of the color analog scale: repeatability of scores in traumatic and nontraumatic injuries. **Academic Emergency Medicine**, Malden, Massachusetts, v. 16, n. 5, p. 465-469, 2009. ISSN 1069-6563. Disponível em: < <http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=105417667&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid> >.
- BURSCH, B. et al. Preliminary validation of a self-efficacy scale for child functioning despite chronic pain (child and parent versions). **Pain**, v. 125, n. 1-2, p. 35-42, 2006. ISSN 0304-3959. Disponível em: < <http://dx.doi.org/10.1016/j.pain.2006.04.026> >.

BUTTNER, W.; FINKE, W. Analysis of behavioural and physiological parameters for the assessment of postoperative analgesic demand in newborns, infants and young children: a comprehensive report on seven consecutive studies. **Paediatr Anaesth**, v. 10, n. 3, p. 303-18, 2000. ISSN 1155-5645.

CASTARLENAS, E. et al. Psychometric Properties of the Numerical Rating Scale to Assess Self-Reported Pain Intensity in Children and Adolescents: A Systematic Review. **Clin J Pain**, v. 33, n. 4, p. 376-383, Apr 2017. ISSN 0749-8047.

CHAN, J. Y.; VON BAEYER, C. L. Cognitive developmental influences on the ability of preschool-aged children to self-report their pain intensity. **Pain**, v. 157, n. 5, p. 997-1001, May 2016. ISSN 0304-3959.

CHEN, X. et al. Reliability and validity of the Pediatric Quality of Life Inventory (PedsQL) Short Form 15 Generic Core Scales in Japan. **Qual Life Res**, v. 16, n. 7, p. 1239-49, Sep 2007. ISSN 0962-9343.

CHIAROTTO, A. et al. Cross-cultural adaptation and validity of the Italian version of the Central Sensitization Inventory. **Musculoskelet Sci Pract**, v. 37, p. 20-28, Oct 2018. ISSN 2468-7812.

CLAAR, R. L.; WALKER, L. S. Functional assessment of pediatric pain patients: Psychometric properties of the Functional Disability Inventory. **Pain**, v. 121, n. 1-2, p. 77-84, 2006. ISSN 0304-3959. Disponível em: < <http://dx.doi.org/10.1016/j.pain.2005.12.002> >.

COFFELT, T. A.; BAUER, B. D.; CARROLL, A. E. Inpatient characteristics of the child admitted with chronic pain. **Pediatrics**, v. 132, n. 2, p. e422-9, Aug 2013. ISSN 0031-4005.

CONNOLLY, M. R. et al. Design and Reporting Characteristics of Clinical Trials of Select Chronic and Recurrent Pediatric Pain Conditions: An Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks Systematic Review. **J Pain**, v. 20, n. 4, p. 394-404, Apr 2019. ISSN 1526-5900.

CORTINA, J. M. What is coefficient alpha? An examination of theory and applications. **Journal of Applied Psychology**, US, v. 78, n. 1, p. 98-104, 1993. ISSN 1939-1854(Electronic),0021-9010.

COSMIN. **COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)** 2019.

CRAIG, K. D. et al. Perceiving pain in others: automatic and controlled mechanisms. **J Pain**, v. 11, n. 2, p. 101-8, Feb 2010. ISSN 1526-5900.

CROMBEZ, G. et al. The child version of the pain catastrophizing scale (PCS-C): A preliminary validation. **Pain**, v. 104, n. 3, p. 639-646, 2003. ISSN 0304-3959. Disponível em: < [http://dx.doi.org/10.1016/S0304-3959\(03\)00121-0](http://dx.doi.org/10.1016/S0304-3959(03)00121-0) >.

CUESTA-VARGAS, A. I. et al. Dimensionality and Reliability of the Central Sensitization Inventory in a Pooled Multicountry Sample. **Journal of Pain**, v. 19, n. 3,

p. 317-329, 2018. ISSN 1528-8447. Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2017.11.006> >.

CUNNINGHAM, N. R. et al. Development and validation of the self-reported PROMIS pediatric pain behavior item bank and short form scale. **Pain**, v. 158, n. 7, p. 1323-1331, 2017. ISSN 1872-6623. Disponível em: < <http://dx.doi.org/10.1097/j.pain.0000000000000914> >.

DA SILVA, F. C.; SANTOS THULER, L. C.; DE LEON-CASASOLA, O. A. Validity and reliability of two pain assessment tools in Brazilian children and adolescents. **Journal of Clinical Nursing (John Wiley & Sons, Inc.)**, v. 20, n. 13-14, p. 1842-1848, 2011. ISSN 0962-1067. Disponível em: < <http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=61195571&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid> >.

DAGFINRUD, H.; HAGEN, K. B.; KVIEN, T. K. Physiotherapy interventions for ankylosing spondylitis. **Cochrane Database of Systematic Reviews**, n. 1, 2008. ISSN 1465-1858. Disponível em: < <http://dx.doi.org/10.1002/14651858.CD002822.pub3> >.

DAVERN, M.; CUMMINS, R.; STOKES, M. **Subjective Wellbeing as an Affective-Cognitive Construct**. 2007. 429-449.

DE VET, H. C. et al. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. **Qual Life Res**, v. 14, n. 5, p. 1203-18; discussion 1219-21, 1223-4, Jun 2005. ISSN 0962-9343 .

DEGOTARDI, P. J. et al. Development and evaluation of a cognitive-behavioral intervention for juvenile fibromyalgia. **J Pediatr Psychol**, v. 31, n. 7, p. 714-23, Aug 2006. ISSN 0146-8693 .

DEKKER, C. et al. Dutch version of the Fear of Pain Questionnaire for adolescents with chronic pain. **Disability and Rehabilitation: An International, Multidisciplinary Journal**, United Kingdom, v. 40, n. 11, p. 1326-1332, 2018. ISSN 1464-5165(Electronic),0963-8288.

DESAPRIYA, E. et al. Vision screening of older drivers for preventing road traffic injuries and fatalities. **Cochrane Database of Systematic Reviews**, n. 2, 2014. ISSN 1465-1858. Disponível em: < <http://dx.doi.org/10.1002/14651858.CD006252.pub4> >.

EBRAHIMPOUR, F. et al. Children's description of pain through drawings and dialogs: A concept analysis. **Nurs Open**, v. 6, n. 2, p. 301-312, Apr 2019. ISSN 2054-1058.

ECCLESTON, C.; MALLESON, P. Managing chronic pain in children and adolescents. We need to address the embarrassing lack of data for this common problem. **BMJ (Clinical research ed.)**, v. 326, n. 7404, p. 1408-1409, 2003. ISSN 1756-1833. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12829527> >

ECCLESTON, C. et al. Systematic review of randomised controlled trials of psychological therapy for chronic pain in children and adolescents, with a subset meta-analysis of pain relief. **Pain**, v. 99, n. 1-2, p. 157-65, Sep 2002. ISSN 0304-3959.

EVANS, J. R. et al. Psychometric Properties of the Pain Stages of Change Questionnaire: New Insights on the Measurement of Readiness to Change in Adolescents, Mothers, and Fathers. **Journal of Pain**, v. 16, n. 7, p. 645-656, 2015. ISSN 1528-8447. Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2015.03.012> >.

FAZENY, B. et al. Barriers in cancer pain management. **Wien Klin Wochenschr**, v. 112, n. 22, p. 978-81, Nov 24 2000. ISSN 0043-5325.

FLOYD, F. J.; WIDAMAN, K. F. Factor analysis in the development and refinement of clinical assessment instruments. **Psychological Assessment**, US, v. 7, n. 3, p. 286-299, 1995. ISSN 1939-134X(Electronic),1040-3590.

FORGERON, P. A. et al. Social functioning and peer relationships in children and adolescents with chronic pain: A systematic review. **Pain Res Manag**, v. 15, n. 1, p. 27-41, Jan-Feb 2010. ISSN 1203-6765.

FRIEDRICHSDORF, S. J. et al. Chronic Pain in Children and Adolescents: Diagnosis and Treatment of Primary Pain Disorders in Head, Abdomen, Muscles and Joints. **Children (Basel)**, v. 3, n. 4, Dec 10 2016. ISSN 2227-9067.

FRIEDRICHSDORF, S. J. et al. Pain Outcomes in a US Children's Hospital: A Prospective Cross-Sectional Survey. **Hosp Pediatr**, v. 5, n. 1, p. 18-26, Jan 2015. ISSN 2154-1663.

FROST, M. H. et al. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? **Value Health**, v. 10 Suppl 2, p. S94-s105, Nov-Dec 2007. ISSN 1098-3015.

GAUVAIN-PIQUARD, A. et al. The development of the DEGR(R): A scale to assess pain in young children with cancer. **Eur J Pain**, v. 3, n. 2, p. 165-176, Jun 1999. ISSN 1090-3801.

GERI, T. et al. Cross-cultural adaptation and validation of the Neck Bournemouth Questionnaire in the Italian population. **Quality of Life Research**, v. 24, n. 3, p. 735-745, 2015. ISSN 1573-2649. Disponível em: < <http://dx.doi.org/10.1007/s11136-014-0806-5> >.

GILSON, A. M.; JORANSON, D. E.; MAURER, M. A. Improving state pain policies: recent progress and continuing opportunities. **CA Cancer J Clin**, v. 57, n. 6, p. 341-53, Nov-Dec 2007. ISSN 0007-9235.

GOFFAUX, P. et al. Sex differences in perceived pain are affected by an anxious brain. **Pain**, v. 152, n. 9, p. 2065-73, Sep 2011. ISSN 0304-3959.

GOODMAN, J. E.; MCGRATH, P. J. The epidemiology of pain in children and adolescents: a review. **Pain**, v. 46, n. 3, p. 247-64, Sep 1991. ISSN 0304-3959.

GOODMAN, R. Psychometric properties of the strengths and difficulties questionnaire. **J Am Acad Child Adolesc Psychiatry**, v. 40, n. 11, p. 1337-45, Nov 2001. ISSN 0890-8567.

GUITE, J. W. et al. Readiness to change in pediatric chronic pain: Initial validation of adolescent and parent versions of the Pain Stages of Change Questionnaire. **Pain**, v. 152, n. 10, p. 2301-2311, 2011. ISSN 0304-3959. Disponível em: < <http://dx.doi.org/10.1016/j.pain.2011.06.019> >.

GUYATT, G. et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. **J Clin Epidemiol**, v. 64, n. 4, p. 383-94, Apr 2011. ISSN 0895-4356.

GUYATT, G. H. et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. **Bmj**, v. 336, n. 7650, p. 924-6, Apr 26 2008. ISSN 0959-8138.

HAINSWORTH, K. R. et al. Development and Preliminary Validation of the Child Activity Limitations Questionnaire: Flexible and Efficient Assessment of Pain-Related Functional Disability. **Journal of Pain**, v. 8, n. 9, p. 746-752, 2007. ISSN 1526-5900. Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2007.05.005> >.

HASSETT, A. L. et al. Reports of chronic pain in childhood and adolescence among patients at a tertiary care pain clinic. **J Pain**, v. 14, n. 11, p. 1390-7, Nov 2013. ISSN 1526-5900.

HAUER, J.; HOUTROW, A. J. Pain Assessment and Treatment in Children With Significant Impairment of the Central Nervous System. **Pediatrics**, v. 139, n. 6, Jun 2017. ISSN 0031-4005.

HEALE, R.; TWYCROSS, A. Validity and reliability in quantitative studies. **Evid Based Nurs**, v. 18, n. 3, p. 66-7, Jul 2015. ISSN 1367-6539.

HECHLER, T. et al. [Pain-related coping strategies in children and adolescents with chronic pain. Validation of a German version of the Paediatric Pain Coping Inventory (PPCI revised)]. **Schmerz**, v. 22, n. 4, p. 442-57, Aug 2008. ISSN 0932-433X.

HELGADOTTIR, H. L. Pain management practices in children after surgery. **J Pediatr Nurs**, v. 15, n. 5, p. 334-40, Oct 2000. ISSN 0882-5963.

HERMANN, C. et al. The assessment of pain coping and pain-related cognitions in children and adolescents: current methods and further development. **J Pain**, v. 8, n. 10, p. 802-13, Oct 2007. ISSN 1526-5900.

HERMANN, C. et al. The impact of chronic pain in children and adolescents: Development and initial validation of a child and parent version of the Pain Experience Questionnaire. **Pain**, v. 135, n. 3, p. 251-61, Apr 2008. ISSN 0304-3959.

HICKS, C. L. et al. The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. **Pain**, v. 93, n. 2, p. 173-83, Aug 2001. ISSN 0304-3959.

HOTOPF, M. et al. Why do children have chronic abdominal pain, and what happens to them when they grow up? Population based cohort study. **Bmj**, v. 316, n. 7139, p. 1196-200, Apr 18 1998. ISSN 0959-8138.

HOTOPF, M. et al. Psychosocial and developmental antecedents of chest pain in young adults. **Psychosom Med**, v. 61, n. 6, p. 861-7, Nov-Dec 1999. ISSN 0033-3174.

HOWARD, R. F. Current status of pain management in children. **Jama**, v. 290, n. 18, p. 2464-9, Nov 12 2003. ISSN 0098-7484.

HUGUET, A.; MIRO, J. The severity of chronic pediatric pain: an epidemiological study. **J Pain**, v. 9, n. 3, p. 226-36, Mar 2008. ISSN 1526-5900.

HUGUET, A.; STINSON, J. N.; MCGRATH, P. J. Measurement of self-reported pain intensity in children and adolescents. **J Psychosom Res**, v. 68, n. 4, p. 329-36, Apr 2010. ISSN 0022-3999.

HURLEY-WALLACE, A. et al. Paediatric pain education for health care professionals. **Pain Rep**, v. 4, n. 1, p. e701, Jan-Feb 2019. ISSN 2471-2531.

IOANNIDIS, J. P. et al. Increasing value and reducing waste in research design, conduct, and analysis. **Lancet**, v. 383, n. 9912, p. 166-75, Jan 11 2014. ISSN 0140-6736.

JACOBSON JR, C. J. et al. Qualitative evaluation of pediatric pain behavior, quality, and intensity item candidates and the PROMIS pain domain framework in children with chronic pain. **The Journal of Pain**, Netherlands, v. 16, n. 12, p. 1243-1255, 2015. ISSN 1528-8447(Electronic),1526-5900.

JOHNSTON, C. et al. Pain in Canadian NICUs: have we improved over the past 12 years? **Clin J Pain**, v. 27, n. 3, p. 225-32, Mar-Apr 2011. ISSN 0749-8047.

JONES, G. T. et al. Are common symptoms in childhood associated with chronic widespread body pain in adulthood? Results from the 1958 British Birth Cohort Study. **Arthritis Rheum**, v. 56, n. 5, p. 1669-75, May 2007. ISSN 0004-3591.

JONGUDOMKARN, D.; AUNGSUPAKORN, N.; CAMFIELD, L. The meanings of pain: A qualitative study of the perspectives of children living with pain in north-eastern Thailand. **Nurs Health Sci**, v. 8, n. 3, p. 156-63, Sep 2006. ISSN 1441-0745.

KASHIKAR-ZUCK, S. Treatment of children with unexplained chronic pain. **Lancet**, v. 367, n. 9508, p. 380-2, Feb 4 2006. ISSN 0140-6736.

KASHIKAR-ZUCK, S. et al. Longitudinal evaluation of patient-reported outcomes measurement information systems measures in pediatric chronic pain. **Pain**, v. 157, n. 2, p. 339-347, 2016. ISSN 1872-6623. Disponível em: <
<http://dx.doi.org/10.1097/j.pain.0000000000000378> >.

KASHIKAR-ZUCK, S. et al. Clinical utility and validity of the Functional Disability Inventory among a multicenter sample of youth with chronic pain. **Pain**, v. 152, n. 7,

p. 1600-1607, 2011. ISSN 0304-3959. Disponível em: <
<http://dx.doi.org/10.1016/j.pain.2011.02.050> >.

KASHIKAR-ZUCK, S. et al. Efficacy of cognitive-behavioral intervention for juvenile primary fibromyalgia syndrome. **J Rheumatol**, v. 32, n. 8, p. 1594-602, Aug 2005. ISSN 0315-162X.

KAVANAGH, T.; WATT-WATSON, J. Paediatric pain education: A call for innovation and change. **Paediatr Child Health**, v. 12, n. 2, p. 97-9, Feb 2007. ISSN 1205-7088.

KERNS, R. D.; TURK, D. C.; RUDY, T. E. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). **Pain**, v. 23, n. 4, p. 345-56, Dec 1985. ISSN 0304-3959.

KESZEI, A. P.; NOVAK, M.; STREINER, D. L. Introduction to health measurement scales. **J Psychosom Res**, v. 68, n. 4, p. 319-23, Apr 2010. ISSN 0022-3999.

KIMBERLIN, C. L.; WINTERSTEIN, A. G. Validity and reliability of measurement instruments used in research. **Am J Health Syst Pharm**, v. 65, n. 23, p. 2276-84, Dec 1 2008. ISSN 1079-2082.

KING, S. et al. The epidemiology of chronic pain in children and adolescents revisited: a systematic review. **Pain**, v. 152, n. 12, p. 2729-38, Dec 2011. ISSN 0304-3959.

KONIJNENBERG, A. Y. et al. Children with unexplained chronic pain: do pediatricians agree regarding the diagnostic approach and presumed primary cause? **Pediatrics**, v. 114, n. 5, p. 1220-6, Nov 2004. ISSN 0031-4005.

KONIJNENBERG, A. Y. et al. Children with unexplained chronic pain: substantial impairment in everyday life. **Arch Dis Child**, v. 90, n. 7, p. 680-6, Jul 2005. ISSN 0003-9888.

KORTESLUOMA, R. L.; NIKKONEN, M.; SERLO, W. "You just have to make the pain go away"--children's experiences of pain management. **Pain Manag Nurs**, v. 9, n. 4, p. 143-9, 149.e1-5, Dec 2008. ISSN 1524-9042.

KORTESLUOMA, R. L.; PUNAMAKI, R. L.; NIKKONEN, M. Hospitalized children drawing their pain: the contents and cognitive and emotional characteristics of pain drawings. **J Child Health Care**, v. 12, n. 4, p. 284-300, Dec 2008. ISSN 1367-4935.

KROHNE, H. W. et al. Das "Erziehungsstil-Inventar" (ESI): Konstruktion, Psychometrische Kennwerte Gültigkeitsstudien. [The "Childrearing Style Inventory" (ESI): Construction psychometric values, validity studies.]. **Diagnostica**, Germany, v. 30, n. 4, p. 299-318, 1984. ISSN 2190-622X(Electronic),0012-1924.

LAIRD, K. T. et al. Validation of the Abdominal Pain Index using a revised scoring method. **J Pediatr Psychol**, v. 40, n. 5, p. 517-25, Jun 2015. ISSN 0146-8693.

LANZI, G. et al. The effectiveness of hospitalization in the treatment of paediatric idiopathic headache patients. **Psychopathology**, v. 40, n. 1, p. 1-7, 2007. ISSN 0254-4962.

LEE, R. R. et al. Chronic Pain Assessments in Children and Adolescents: A Systematic Literature Review of the Selection, Administration, Interpretation, and Reporting of Unidimensional Pain Intensity Scales. **Pain Res Manag**, v. 2017, p. 7603758, 2017. ISSN 1203-6765.

LEWANDOWSKI, A. S. et al. Temporal daily associations between pain and sleep in adolescents with chronic pain versus healthy adolescents. **Pain**, v. 151, n. 1, p. 220-5, Oct 2010. ISSN 0304-3959.

LINDEGREN, M. L. et al. Integration of HIV/AIDS services with maternal, neonatal and child health, nutrition, and family planning services. **Cochrane Database of Systematic Reviews**, n. 9, 2012. ISSN 1465-1858. Disponível em: < <http://dx.doi.org/10.1002/14651858.CD010119> >.

LIOSSI, C. et al. The effectiveness of online pain resources for health professionals: a systematic review with subset meta-analysis of educational intervention studies. **Pain**, v. 159, n. 4, p. 631-643, Apr 2018. ISSN 0304-3959.

LIOSSI, C. et al. Effectiveness of interdisciplinary interventions in paediatric chronic pain management: a systematic review and subset meta-analysis. **Br J Anaesth**, Feb 28 2019. ISSN 0007-0912.

MANWORREN, R. C.; STINSON, J. Pediatric Pain Measurement, Assessment, and Evaluation. **Semin Pediatr Neurol**, v. 23, n. 3, p. 189-200, Aug 2016. ISSN 1071-9091.

MAREC-BERARD, P. et al. HEDEN Pain Scale: A Shortened Behavioral Scale for Assessment of Prolonged Cancer or Postsurgical Pain in Children Aged 2 to 6 Years. **Pediatr Hematol Oncol**, v. 32, n. 5, p. 291-303, 2015. ISSN 0888-0018.

MARTIN, S. et al. Development and validation of the english pain interference index and pain interference index-parent report. **Pain Medicine**, v. 16, n. 2, p. 367-373, 2015. ISSN 1526-4637.

MCCRACKEN, L. M.; GAUNTLETT-GILBERT, J.; ECCLESTON, C. Acceptance of pain in adolescents with chronic pain: Validation of an adapted assessment instrument and preliminary correlation analyses. **European Journal of Pain**, v. 14, n. 3, p. 316-320, 2010. ISSN 1090-3801. Disponível em: < <http://dx.doi.org/10.1016/j.ejpain.2009.05.002> >.

MCGRATH, P. A. et al. A new analogue scale for assessing children's pain: an initial validation study. **Pain**, v. 64, n. 3, p. 435-43, Mar 1996. ISSN 0304-3959.

MILLER, M. M. et al. Perceived Injustice Is Associated With Pain and Functional Outcomes in Children and Adolescents With Chronic Pain: A Preliminary Examination. **Journal of Pain**, v. 17, n. 11, p. 1217-1226, 2016. ISSN 1528-8447. Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2016.08.002> >.

MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **PLoS Med**, v. 6, n. 7, p. e1000097, Jul 21 2009. ISSN 1549-1277.

MOKKINK, L. B. et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. **Qual Life Res**, v. 27, n. 5, p. 1171-1179, May 2018. ISSN 0962-9343.

MOKKINK, L. B. et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. **Braz J Phys Ther**, v. 20, n. 2, p. 105-13, Jan 19 2016. ISSN 1413-3555.

MOKKINK, L. B. et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. **J Clin Epidemiol**, v. 63, n. 7, p. 737-45, Jul 2010. ISSN 0895-4356.

MUNN, Z. et al. The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI). **Int J Evid Based Healthc**, v. 17, n. 1, p. 36-43, Mar 2019. ISSN 1744-1595.

OFFENBÄCHER, M. et al. Functional limitations in children and adolescents suffering from chronic pain: validation and psychometric properties of the German Functional Disability Inventory (FDI-G). **Rheumatology International**, v. 36, n. 10, p. 1439-1448, 2016. ISSN 1437-160X. Disponível em: < <http://dx.doi.org/10.1007/s00296-016-3504-5> >.

PALERMO, T. et al. **Special considerations in conducting clinical trials of chronic pain management interventions in children and adolescents and their families.** 2018. 1.

PALERMO, T. M. et al. Validation of a self-report questionnaire version of the Child Activity Limitations Interview (CALI): The CALI-21. **Pain**, Netherlands, v. 139, n. 3, p. 644-652, 2008. ISSN 1872-6623(Electronic),0304-3959.

PALERMO, T. M. et al. Waiting for a Pediatric Chronic Pain Clinic Evaluation: A Prospective Study Characterizing Waiting Times and Symptom Trajectories. **J Pain**, v. 20, n. 3, p. 339-347, Mar 2019. ISSN 1526-5900.

PALERMO, T. M. et al. Development and validation of the Child Activity Limitations Interview: A measure of pain-related functional impairment in school-age children and adolescents. **Pain**, v. 109, n. 3, p. 461-470, 2004. ISSN 0304-3959. Disponível em: < <http://dx.doi.org/10.1016/j.pain.2004.02.023> >.

PERGOLIZZI, J. V., JR.; RAFFA, R. B.; TAYLOR, R., JR. Treating acute pain in light of the chronification of pain. **Pain Manag Nurs**, v. 15, n. 1, p. 380-90, Mar 2014. ISSN 1524-9042.

PERQUIN, C. W. et al. The natural course of chronic benign pain in childhood and adolescence: a two-year population-based follow-up study. **Eur J Pain**, v. 7, n. 6, p. 551-9, 2003. ISSN 1090-3801.

PIELECH, M. et al. Pain catastrophizing in children with chronic pain and their parents: Proposed clinical reference points and reexamination of the PCS measure.

Pain, v. 155, n. 11, p. 2360-2367, 2014. ISSN 1872-6623. Disponível em: < <http://dx.doi.org/10.1016/j.pain.2014.08.035> >.

POLIT, D. F. Assessing measurement in health: Beyond reliability and validity. **Int J Nurs Stud**, v. 52, n. 11, p. 1746-53, Nov 2015. ISSN 0020-7489.

POLIT, D. F.; BECK, C. T. The content validity index: are you sure you know what's being reported? Critique and recommendations. **Res Nurs Health**, v. 29, n. 5, p. 489-97, Oct 2006. ISSN 0160-6891.

PRINSEN, C. A. C. et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. **Qual Life Res**, v. 27, n. 5, p. 1147-1157, May 2018. ISSN 0962-9343.

REID, G. J.; GILBERT, C. A.; MCGRATH, P. J. The pain coping questionnaire: Preliminary validation. **Pain**, v. 76, n. 1-2, p. 83-96, 1998. ISSN 0304-3959. Disponível em: < [http://dx.doi.org/10.1016/S0304-3959\(98\)00029-3](http://dx.doi.org/10.1016/S0304-3959(98)00029-3) >.

ROSSMANN, P. DTK-Depressions test für Kinder [DTC-Depression Test for Children]. **Huber Verlag, Bern, Switzerland**, 2005.

ROTH, R. S.; PUNCH, M. R.; BACHMAN, J. E. Patient beliefs about pain diagnosis in chronic pelvic pain: relation to pain experience, mood and disability. **J Reprod Med**, v. 56, n. 3-4, p. 123-9, Mar-Apr 2011. ISSN 0024-7758.

SALLFORS, C.; FASTH, A.; HALLBERG, L. R. Oscillating between hope and despair--a qualitative study. **Child Care Health Dev**, v. 28, n. 6, p. 495-505, Nov 2002. ISSN 0305-1862.

SALMOND, S. S. Evaluating the reliability and validity of measurement instruments. **Orthop Nurs**, v. 27, n. 1, p. 28-30, Jan-Feb 2008. ISSN 0744-6020.

SANYA, E. O.; KOLO, P. M.; MAKUSIDI, M. A. A survey on doctors' knowledge and attitude of treating chronic pain in three tertiary hospitals in Nigeria. **Niger Med J**, v. 55, n. 2, p. 106-10, Mar 2014. ISSN 0300-1652.

SAUL, R.; PETERS, J.; BRUCE, E. Assessing acute and chronic pain in children and young people. **Nurs Stand**, v. 31, n. 10, p. 51-63, Nov 2 2016. ISSN 0029-6570.

SCHECHTER, N. L. Functional pain: time for a new name. **JAMA Pediatr**, v. 168, n. 8, p. 693-4, Aug 2014. ISSN 2168-6203.

SCHNEEWIND, K. Familienpsychologie (2. Aufl.) Stuttgart: Kohlhammer. 1999.

SIL, S.; COHEN, L. L.; DAMPIER, C. Pediatric pain screening identifies youth at risk of chronic pain in sickle cell disease. **Pediatric Blood and Cancer**, v. 66, n. 3, 2019. ISSN 1545-5017. Disponível em: < <http://dx.doi.org/10.1002/pbc.27538> >.

SIMONS, L. E. et al. The Fear of Pain Questionnaire (FOPQ): Assessment of pain-related fear among children and adolescents with chronic pain. **Journal of Pain**, v. 12, n. 6, p. 677-686, 2011. ISSN 1526-5900. Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2010.12.008> >.

SIMONS, L. E. et al. Pediatric pain screening tool: Rapid identification of risk in youth with pain complaints. **Pain**, v. 156, n. 8, p. 1511-1518, 2015. ISSN 1872-6623 Disponível em: < <http://dx.doi.org/10.1097/j.pain.0000000000000199> >.

SNYDER, C. F. et al. Patient-reported outcome instrument selection: designing a measurement strategy. **Value Health**, v. 10 Suppl 2, p. S76-85, Nov-Dec 2007. ISSN 1098-3015.

STAHLSCHMIDT, L. et al. Assessment of Pain-related Disability in Pediatric Chronic Pain. **Clinical Journal of Pain**, v. 34, n. 12, p. 1173-1179, 2018. ISSN 1536-5409 Disponível em: < <http://dx.doi.org/10.1097/AJP.0000000000000646> >.

STANFORD, E. A. et al. The frequency, trajectories and predictors of adolescent recurrent pain: a population-based approach. **Pain**, v. 138, n. 1, p. 11-21, Aug 15 2008. ISSN 0304-3959.

STEVENS, B. J. et al. Epidemiology and management of painful procedures in children in Canadian hospitals. **Cmaj**, v. 183, n. 7, p. E403-10, Apr 19 2011. ISSN 0820-3946.

STEVENS, B. J. et al. Pain assessment and intensity in hospitalized children in Canada. **J Pain**, v. 13, n. 9, p. 857-65, Sep 2012. ISSN 1526-5900.

STINSON, J. N. et al. Systematic review of the psychometric properties, interpretability and feasibility of self-report pain intensity measures for use in clinical trials in children and adolescents. **Pain**, v. 125, n. 1-2, p. 143-57, Nov 2006. ISSN 0304-3959.

STONE, A. L. et al. Pediatric Pain Beliefs Questionnaire: Psychometric Properties of the Short Form. **Journal of Pain**, v. 17, n. 9, p. 1036-1044, 2016. ISSN 1528-8447 Disponível em: < <http://dx.doi.org/10.1016/j.jpain.2016.06.006> >.

STREINER, D. L. Figuring out factors: the use and misuse of factor analysis. **Can J Psychiatry**, v. 39, n. 3, p. 135-40, Apr 1994. ISSN 0706-7437.

STREINER, D. L.; KOTTNER, J. Recommendations for reporting the results of studies of instrument and scale development and testing. n. 1365-2648 (Electronic),

TAYLOR, E. M.; BOYER, K.; CAMPBELL, F. A. Pain in hospitalized children: a prospective cross-sectional survey of pain prevalence, intensity, assessment and management in a Canadian pediatric teaching hospital. **Pain Res Manag**, v. 13, n. 1, p. 25-32, Jan-Feb 2008. ISSN 1203-6765.

TERWEE, C. B. et al. Quality criteria were proposed for measurement properties of health status questionnaires. **J Clin Epidemiol**, v. 60, n. 1, p. 34-42, Jan 2007. ISSN 0895-4356.

TERWEE, C. B. et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. **Qual Life Res**, v. 21, n. 4, p. 651-7, May 2012. ISSN 0962-9343.

TERWEE, C. B. et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. **Qual Life Res**, v. 27, n. 5, p. 1159-1170, May 2018. ISSN 0962-9343.

THURNER, F.; TEWES, U. **Der Kinder-Angst-Test II:(KAT II)**. Hogrefe, 2000.

TSZE, D. S. et al. Validation of self-report pain scales in children. **Pediatrics**, v. 132, n. 4, p. e971-e979, 2013. ISSN 1098-4275 Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/23999954>

TUMIN, D. et al. Health Care Utilization and Costs Associated With Pediatric Chronic Pain. **J Pain**, v. 19, n. 9, p. 973-982, Sep 2018. ISSN 1526-5900.

VAN DE VOORDE, P. et al. The long-term outcome after severe trauma of children in Flanders (Belgium): A population-based cohort study using the International Classification of Functioning-related outcome score. **European Journal of Pediatrics**, v. 170, n. 1, p. 65-73, 2011. ISSN 0340-6199. Disponível em: <
<http://dx.doi.org/10.1007/s00431-010-1271-y> >.

VON BAEYER, C. L. et al. Developmental Data Supporting Simplification of Self-Report Pain Scales for Preschool-Age Children. **The Journal of Pain**, v. 14, n. 10, p. 1116-1121, 2013/10/01/ 2013. ISSN 1526-5900. Disponível em: <
<http://www.sciencedirect.com/science/article/pii/S1526590013009711> >.

VON BAEYER, C. L.; SPAGRUD, L. J. Systematic review of observational (behavioral) measures of pain for children and adolescents aged 3 to 18 years. **Pain**, v. 127, n. 1-2, p. 140-50, Jan 2007. ISSN 0304-3959.

WAGER, J. et al. Classifying the severity of paediatric chronic pain—An application of the chronic pain grading. **European Journal of Pain**, United Kingdom, v. 17, n. 9, p. 1393-1402, 2013. ISSN 1532-2149(Electronic),1090-3801.

WALKER, L. S. et al. A typology of pain coping strategies in pediatric patients with chronic abdominal pain. **Pain**, v. 137, n. 2, p. 266-75, Jul 15 2008. ISSN 0304-3959.

WALKER, L. S. et al. Children's Somatization Inventory: psychometric properties of the revised form (CSI-24). **Journal of pediatric psychology**, v. 34, n. 4, p. 430-440, 2009. ISSN 1465-735X. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/18782857>

WALKER, L. S. et al. Functional abdominal pain in childhood and adolescence increases risk for chronic pain in adulthood. **Pain**, v. 150, n. 3, p. 568-72, Sep 2010. ISSN 0304-3959.

WALKER, L. S. et al. Testing a model of pain appraisal and coping in children with chronic abdominal pain. **Health Psychology**, Washington, District of Columbia, v. 24, n. 4, p. 364-374, 2005. ISSN 0278-6133. Disponível em: <
<http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=106530460&lang=pt-br&site=ehost-live&authtype=ip,cookie,uid> >.

WALTON, M. K. et al. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for

Outcomes Research Task Force. **Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research**, v. 18, n. 6, p. 741-752, 2015. ISSN 1524-4733. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26409600> >

WARLOW, T. A.; HAIN, R. D. W. 'Total Pain' in Children with Severe Neurological Impairment. **Children (Basel)**, v. 5, n. 1, Jan 18 2018. ISSN 2227-9067.

WHO Guidelines Approved by the Guidelines Review Committee. In: (Ed.). **WHO Guidelines on the Pharmacological Treatment of Persisting Pain in Children with Medical Illnesses**. Geneva: World Health Organization., 2012.

WILLIS, M. H. et al. FLACC Behavioral Pain Assessment Scale: a comparison with the child's self-report. **Pediatr Nurs**, v. 29, n. 3, p. 195-8, May-Jun 2003. ISSN 0097-9805 .

WONG, D. L.; BAKER, C. M. Pain in children: comparison of assessment scales. **Pediatr Nurs**, v. 14, n. 1, p. 9-17, Jan-Feb 1988. ISSN 0097-9805.

WOODS, K.; OSTROWSKI-DELAHANTY, S. Psychometric Properties of the Psychosocial Assessment Tool-Chronic Pain Version in Families of Children with Headache. **Journal of Child Neurology**, v. 32, n. 8, p. 766-773, 2017. ISSN 1708-8283. Disponível em: < <http://dx.doi.org/10.1177/0883073817707111> >.

ZAMANZADEH, V. et al. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. **J Caring Sci**, v. 4, n. 2, p. 165-78, Jun 2015. ISSN 2251-9920.

ZHU, L. M. et al. Improvements in pain outcomes in a Canadian pediatric teaching hospital following implementation of a multifaceted knowledge translation initiative. **Pain Res Manag**, v. 17, n. 3, p. 173-9, May-Jun 2012. ISSN 1203-6765.

APÊNDICE I

Protocolo de revisão sistemática submetido ao periódico científico Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports (JBISRIIR).

JBIR Database of Systematic Reviews and Implementation Reports
Chronic pain assessment tools in hospitalized children: a systematic review protocol
--Manuscript Draft--

Manuscript Number:	
Keywords:	assessments; children; chronic pain; inpatients.
Corresponding Author:	Bruna de Moraes Rubim Alelaf, MSc Federal University of Piauí Parnaíba, Piauí BRAZIL
Corresponding Author's Institution:	Federal University of Piauí
Corresponding Author E-Mail:	brunamoraesrubim@hotmail.com
Manuscript Region of Origin:	BRAZIL
First Author:	Bruna de Moraes Rubim Alelaf
Other Authors:	Emerson Diogenes de Medeiros, PhD
	Abrahão Fontes Baptista, PhD
	Katia Nunes Sá, PhD
	Fuad Ahmad Hazime, PhD

Chronic pain assessment tools in hospitalized children: a systematic review protocol of measurement properties

Bruna de Moraes Rubim Alelaf¹

Emerson Diogenes de Medeiros²

Abrahão Fontes Baptista³

Katia Nunes Sá⁴

Fuad Ahmad Hazime¹

¹Biomedical Master Science Program, Federal Univeristy of Piau , Parna ba, Brazil.

²Psychology Master Science Program, Federal Univeristy of Piau , Parna ba, Brazil.

³Center for Mathematics, Computation and Cognition, Federal Univeristy of ABC, S o Bernardo do Campo, Brazil.

⁴University of Salvador and Bahia School of Medicine and Public Health, Salvador – Brazil.

Acknowledgments

This review will contribute to a Masters in Biomedical Sciences at Federal University of Piau  for BMRA.

Conflicts of interest

The authors have no conflicts of interest to declare.

Corresponding author

Bruna de Moraes Rubim Alelaf

+55 (86) 3323-5209

Biomedical Master Science Program, Federal University of Piau , Parna ba, Brasil. S o Sebast o Avenue, 2819, Parna ba, Piau , Brazil. Zip code: 64202-020.

Introduction

Pediatric chronic pain is a challenging condition that affects approximately 20% to 35% of children around the world.(Goodman e McGrath, 1991; Stanford *et al.*, 2008; King *et al.*, 2011) However, despite its prevalence, a significant proportion of hospitalized children receive inadequate assessment and pain management,(Helgadottir, 2000; Taylor *et al.*, 2008; Johnston *et al.*, 2011; Stevens *et al.*, 2011; Stevens *et al.*, 2012; Friedrichsdorf *et al.*, 2015) which renders treatment difficult and may lead to physiological and psychological alterations and adverse emotional effects.(Anand e Scalzo, 2000; Holsti *et al.*, 2004; Grunau *et al.*, 2006; Holsti *et al.*, 2007; Matziou *et al.*, 2016) In addition, persistent pain during childhood enhances the probability of the continuance of chronic pain into adulthood.(Hassett *et al.*, 2013) Another consequence of pediatric chronic pain is its high costs for society and health services. In the United States, for instance, US\$19.5 billion is spent annually to treat children and teenagers with chronic pain.(Groenewald *et al.*, 2014) Intensive treatment is also necessary in some cases,(Hechler *et al.*, 2010) which increases the estimated costs.(Groenewald *et al.*, 2014)

New concepts of pain chronification have raised new possibilities in terms of pain management, but clinicians and researchers still struggle with pain assessment, especially in hospitalized children. These difficulties may be associated with inadequate instruments, the insufficient knowledge of health professionals,(Taylor *et al.*, 2008; Franck e Bruce, 2009) or the lack of pain assessment protocols.(Taylor *et al.*, 2008; Walther-Larsen *et al.*, 2017) Moreover, lack of clear information about evaluation processes and application of instruments during clinical practice make it difficult to interpret the efficacy of pain management strategies.(Lee *et al.*, 2017; Walther-Larsen *et al.*, 2017)

Many specific tools have been proposed to describe and quantify pain sensation, such as the Wong-Baker FACES Scale and the Numerical Rating Scale (NRS),(Saul *et al.*, 2016b) but self-reporting is considered the gold standard.(Saul *et al.*, 2016b; Speer *et al.*, 2016) However, for children with limited cognitive and verbal skills, behavioral observation instruments are more appropriate, such as the Face, Legs, Activity, Cry, Consolability (FLACC) scale.(Merkel *et al.*, 1997; Speer *et al.*, 2016) As hospitalized children with persistent pain may have difficulties expressing their pain sensations due to stress, fear, or anxiety, an ideal evaluation should comprise self-report and behavioral assessments.(Von Baeyer, 2006; Speer *et al.*, 2016) Other factors should also be taken into account, such as the level of neuro-psychomotor development and children's previous pain condition.(Merkel e Malviya, 2000)

Although more chronic pain assessment instruments are needed for clinical and scientific purposes, they must be reliable, validated, and effective.(Manworren e Stinson, 2016) According to our previous review in PROSPERO, MEDLINE, the Cochrane Database of Systematic Reviews, and the JBI Database of Systematic Reviews and Implementation Reports, we found only one systematic review that investigated pain instruments in hospitalized children. However, this particular systematic review protocol aimed to analyze acute pain instruments. The identification of precise instruments to evaluate chronic pain in hospitalized children, with acceptable reliability and validity, is of great interest for parents, caregivers, clinicians, and researchers. That is, the more precise the pain assessment, the greater the probability of achieving treatment efficacy. This current review aims to systematically summarize and compare the quality of the measurement properties of all chronic pain assessment instruments for evaluation of hospitalized children.

Review question/objective: The aim of this review is to evaluate and summarize the quality of the measurement properties of chronic pain assessment tools that are used in hospitalized children.

Keywords

assessments; children; chronic pain; inpatients.

Inclusion criteria

Participants

Studies will be considered for inclusion in the review if the participants were hospitalized children between the ages of two and 12 years, with a medical diagnose of chronic pain, independently of its origin or complaints of persistent (ongoing) or recurrent (episodic) pain lasting more than three months.

Construct

This review will consider pain assessment instruments used to evaluate chronic pain in hospitalized children.

Instruments

Assessment tools such as questionnaires, scales, inventories or indexes used to investigate chronic pain conditions in hospitalized children will be included.

Outcomes

Studies that include, at least, one measurement properties as outcomes will be included. The following measurement properties will be investigated:

Reliability: internal consistency, test-retest reliability, inter-rater and intra-rater reliability.

Validity: content validity, face validity, construct validity, structural validity, cross-cultural validity, hypotheses testing, criterion validity, responsiveness, sensitivity to change.

Each measurement property will be rated according with the criteria for good measurement properties.(Prinsen *et al.*, 2018)

Types of studies

This review will consider only instrument evaluation studies.

Methods

The current review will be conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) recommendations(Moher *et al.*, 2009) and was previously registered with the International Prospective Register of Systematic Reviews (PROSPERO) under n. CRD42018099021.

Search strategy

The search strategy will aim to identify all studies from inception to present with no limitation of language. Titles and abstracts will be reviewed considering free text words and index terms of our four key elements: construct (Improve Chronic Pain); population (hospitalized children); type of instruments (pain assessment tools) and measurement properties (validity, reliability and generalizability). The databases to be searched include: PubMed, EMBASE, CINAHL, Web of Science, PsycNET, Cochrane Database

for Systematic Reviews, and Google Scholar. Search terms for measurement properties, available for Pubmed and EMBASE, will be performed using sensitive validated search filter in accordance with COSMIN website.(Klassen *et al.*, 2006) Reference lists of all identified reports and articles will also be searched for additional studies. The full search strategy for the PubMed database is detailed in Appendix I.

Study selection

Following the search, all identified citations will be collated and uploaded into EndNote version X6 (Clarivate Analytics, PA, USA) and duplicates removed. Two independent investigators will assess the selected studies for methodological quality prior to inclusion in the review, using standardized critical appraisal instruments from the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI) (Joanna Briggs Institute, Adelaide, Australia). A third investigator will arbitrate disagreements.

Assessment of methodological quality

The quality of the selected articles will be appraised, scored, and qualified using the checklist risk of bias of the COSMIN.(Mokkink *et al.*, 2018) The COSMIN checklist is a standardized tool for evaluating the methodological quality of studies based on measurement properties, such as internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity, and responsiveness.

Data extraction and synthesis

Data will be extracted from papers that are included in the review using a standardized data extraction tool from the JBI SUMARI and adapted to the specific elements of a psychometric review. Extracted data will include characteristics of pain assessment tools, characteristics of the include sample and results on measurement properties. If data are missing or unclear, authors will contact the investigators for clarification. Disagreements between the two independent reviewers will be arbitrated by a third investigator. Subsequently, a detailed analysis will be offered in the discussion section.

References

1. Goodman, J.E. and P.J. McGrath, The epidemiology of pain in children and adolescents: a review. *Pain*, 1991. 46(3): p. 247-64.
2. King, S., et al., The epidemiology of chronic pain in children and adolescents revisited: a systematic review. *Pain*, 2011. 152(12): p. 2729-38.
3. Stanford, E.A., C.T. Chambers, J.C. Biesanz, and E. Chen, The frequency, trajectories and predictors of adolescent recurrent pain: a population-based approach. *Pain*, 2008. 138(1): p. 11-21.
4. Helgadottir, H.L., Pain management practices in children after surgery. *J Pediatr Nurs*, 2000. 15(5): p. 334-40.
5. Johnston, C., K.J. Barrington, A. Taddio, R. Carbajal, and F. Filion, Pain in Canadian NICUs: have we improved over the past 12 years? *Clin J Pain*, 2011. 27(3): p. 225-32.
6. Stevens, B.J., et al., Epidemiology and management of painful procedures in children in Canadian hospitals. *Cmaj*, 2011. 183(7): p. E403-10.
7. Stevens, B.J., et al., Pain assessment and intensity in hospitalized children in Canada. *J Pain*, 2012. 13(9): p. 857-65.
8. Taylor, E.M., K. Boyer, and F.A. Campbell, Pain in hospitalized children: a prospective cross-sectional survey of pain prevalence, intensity, assessment and management in a Canadian pediatric teaching hospital. *Pain Res Manag*, 2008. 13(1): p. 25-32.
9. Friedrichsdorf, S.J., et al., Pain Outcomes in a US Children's Hospital: A Prospective Cross-Sectional Survey. *Hosp Pediatr*, 2015. 5(1): p. 18-26.
10. Anand, K.J. and F.M. Scalzo, Can adverse neonatal experiences alter brain development and subsequent behavior? *Biol Neonate*, 2000. 77(2): p. 69-82.
11. Grunau, R.E., L. Holsti, and J.W. Peters, Long-term consequences of pain in human neonates. *Semin Fetal Neonatal Med*, 2006. 11(4): p. 268-75.
12. Holsti, L., R.E. Grunau, T.F. Oberlander, and M.F. Whitfield, Specific Newborn Individualized Developmental Care and Assessment Program movements are associated with acute pain in preterm infants in the neonatal intensive care unit. *Pediatrics*, 2004. 114(1): p. 65-72.
13. Holsti, L., J. Weinberg, M.F. Whitfield, and R.E. Grunau, Relationships between adrenocorticotrophic hormone and cortisol are altered during clustered nursing care in preterm infants born at extremely low gestational age. *Early Hum Dev*, 2007. 83(5): p. 341-8.
14. Matziou, V., et al., Perceptions of children and their parents about the pain experienced during their hospitalization and its impact on parents' quality of life. *Jpn J Clin Oncol*, 2016. 46(9): p. 862-70.
15. Hassett, A.L., et al., Reports of chronic pain in childhood and adolescence among patients at a tertiary care pain clinic. *J Pain*, 2013. 14(11): p. 1390-7.
16. Groenewald, C.B., B.S. Essner, D. Wright, M.D. Fesinmeyer, and T.M. Palermo, The economic costs of chronic pain among a cohort of treatment-seeking adolescents in the United States. *J Pain*, 2014. 15(9): p. 925-33.

17. Hechler, T., M. Dobe, and B. Zernikow, Commentary: A worldwide call for multimodal inpatient treatment for children and adolescents suffering from chronic pain and pain-related disability. *J Pediatr Psychol*, 2010. 35(2): p. 138-40.
18. Franck, L.S. and E. Bruce, Putting pain assessment into practice: why is it so painful? *Pain Res Manag*, 2009. 14(1): p. 13-20.
19. Walther-Larsen, S., et al., Pain prevalence in hospitalized children: a prospective cross-sectional survey in four Danish university hospitals. *Acta Anaesthesiol Scand*, 2017. 61(3): p. 328-337.
20. Lee, R.R., A. Rashid, D. Ghio, W. Thomson, and L. Cordingley, Chronic Pain Assessments in Children and Adolescents: A systematic Literature Review of the Selection, Administration, Interpretation, and Reporting of Unidimensional Pain Intensity Scales. *Pain Res Manag*, 2017. 2017: p. 7603758.
21. Saul, R., J. Peters, and E. Bruce, Assessing acute and chronic pain in children and young people. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, 2016. 31(10): p. 51-63.
22. Speer, K., T. Chamblee, and J. Tidwell, An evaluation of instruments for identifying acute pain among hospitalized pediatric patients: a systematic review protocol. *JBI Database System Rev Implement Rep*, 2016. 13(12): p. 25-36.
23. Merkel, S.I., T. Voepel-Lewis, J.R. Shayevitz, and S. Malviya, The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatr Nurs*, 1997. 23(3): p. 293-7.
24. von Baeyer, C.L., Children's self-reports of pain intensity: scale selection, limitations and interpretation. *Pain Res Manag*, 2006. 11(3): p. 157-62.
25. Merkel, S. and S. Malviya, Pediatric pain, tools, and assessment. *J Perianesth Nurs*, 2000. 15(6): p. 408-14.
26. Manworren, R.C. and J. Stinson, Pediatric Pain Measurement, Assessment, and Evaluation. *Semin Pediatr Neurol*, 2016. 23(3): p. 189-200.
27. Prinsen, C.A.C., et al., COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*, 2018. 27(5): p. 1147-1157.
28. Moher, D., A. Liberati, J. Tetzlaff, and D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*, 2009. 6(7): p. e1000097.
29. *COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)*. Available from: <http://www.cosmin.nl>.
30. Mokkink, L.B., et al., COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*, 2018. 27(5): p. 1171-1179.

APÊNDICE II

Estratégia de busca

Base de dados PubMed

1.	"chronic pain" [MeSH Terms] OR "chronic pain" [Text Word] OR "Chronic Pains" OR "Pains, Chronic" OR "Pain, Chronic" OR "Widespread Chronic Pain" OR "Chronic Pain, Widespread" OR "Chronic Pains, Widespread" OR "Pain, Widespread Chronic" OR "Pains, Widespread Chronic" OR "Widespread Chronic Pains"
2.	"child" [MeSH Terms] OR "child" [Text Word] OR "children" [Text Word] OR "child, preschool"[MeSH Terms] OR "Preschool Child" OR "Children, Preschool" OR "Preschool Children"
3.	"intensive care units" [MeSH Terms] OR "intensive care unit" [Text Word] OR "hospitals" [MeSH Terms] OR "hospitals" [Text Word] OR "hospitalization" [MeSH Terms] OR "hospitalization" [Text Word] OR "nurseries" [MeSH Terms] OR "nurseries" [Text Word] OR "pediatrics" [MeSH Terms] OR "pediatrics" [Text Word]
4.	"weights and measures" [MeSH Terms] OR "scale" [Text Word] OR "personality inventory" [MeSH Terms] OR "personality inventory" [Text Word] OR "assessment" [All Fields] OR "measure" [Text Word] OR "checklist" [MeSH Terms] OR "checklist" [Text Word] OR "abstracting and indexing as topic" [MeSH Terms] OR "abstracting and indexing as topic" [Text Word] OR "index" [Text Word] OR "score" [All Fields] OR "surveys and questionnaires" [MeSH Terms] OR "questionnaire"[Text Word]
5.	"psychometrics" [MeSH Terms] OR "psychometrics" [Text Word] OR "reproducibility of results" [MeSH Terms] OR reliability and validity [Text Word] OR "internal consistency" [Text Word] OR "data accuracy" [MeSH Terms] OR data accuracy [Text Word] OR "Responsiveness"[Text Word]
1 AND 2 AND 3 AND 4 AND 5	

APÊNDICE III

Instrumento para extração de dados

Características do estudo

Título	
Ano de publicação	
Autor	
Desenho metodológico	
Objetivos	
Cenário	

Características da amostra de inclusão

Idade	
Gênero	
Características clínicas dos participantes	
Tamanho da amostra	

Características das ferramentas de avaliação

Ferramenta que se pretende avaliar propriedades de medida	
Modo de administração	
Avaliador	
Respondente	
O avaliador passou por treinamento antes de aplicar o instrumento de avaliação?	

Houve avaliação da capacidade cognitiva da criança? (Casos de autorrelato)	
--	--

Propriedades de medição

Validade de conteúdo		Erro de medida	
Validade estrutural		Validade de critério	
Consistência interna		Validade de construto	
Validade transcultural		Responsividade	
Confiabilidade			
Ferramentas aplicadas para auxiliar no processo de validação			

Descrição dos resultados

COSMIN Risk of Bias checklist

Date: July, 2018



Contact

L.B. Mokkink, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
Amsterdam Public Health research institute
P.O. box 7057
1007 MB Amsterdam
The Netherlands
Website: www.cosmin.nl
E-mail: w.mokkink@vumc.nl

How to cite the COSMIN Risk of Bias Checklist

Please refer to the following studies when using the COSMIN Risk of Bias Checklist:

Mokkink, L.B., De Vet, H.C.W., Prinsen, C.A.C, Patrick, D.L., Alonso, J., Bouter, L.M., et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Accepted for publication in Quality of Life Research.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., Vet, H. C., et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures.

Submitted.

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Vet, H. C., Westerman, M. J., Patrick, D. L., et al. COSMIN methodology for evaluating the content validity of Patient-Reported Outcome Measures: a Delphi study. Submitted.

For details on how to use the COSMIN risk of Bias checklist see 'COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) – user manual' and 'COSMIN methodology for assessing the content validity of Patient-Reported Outcome Measures (PROMs) - user manual' available from our website www.cosmin.nl.

Abbreviations used:

CTT – classical test theory; DIF – differential item functioning; IRT – Item response theory; MGCFA – multi-group confirmatory factor analysis; MI – measurement invariance; NA – not applicable; PROM – patient-reported outcome measure; 1PL model – 1 parameter IRT model; 2PL model – 2 parameter IRT model

Instructions

Tick the boxes that need to be completed for the article

	COSMIN Risk of Bias checklist
	Box 1. PROM development
	Box 2. Content validity
	Box 3. Structural validity
	Box 4. Internal consistency
	Box 5. Cross-cultural validity\Measurement invariance
	Box 6. Reliability
	Box 7. Measurement error

	Box 8. Criterion validity
	Box 9. Hypotheses testing for construct validity
	Box 10. Responsiveness

To assess the methodological quality of each study, i.e. assessing the risk of bias of the result of a study, the corresponding COSMIN Risk of Bias box should be completed. To determine the overall quality of a study the lowest rating of any standard in the box is taken (i.e. “the worst score counts” principle). For example, if for a reliability study one item in a box is rated as ‘inadequate’, the overall methodological quality of that reliability study is rated as ‘inadequate’. The response option ‘NA’ (not applicable) is at issue for some standards. For example, when a study on structural validity is based on CTT, the standard on IRT is not applicable and this standard should not be considered in the “worst score counts”-rating for that specific study. For standards where this option is not at issue, these cells are grey and shouldn’t be used.

Box 1. PROM development

1a. PROM design

General design requirements

	very good	adequate	doubtful	inadequate	NA
1 Is a clear description provided of the construct to be measured?	Construct clearly described			Construct not clearly described	
2 Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?	Origin of the construct clear		Origin of the construct not clear		
3 Is a clear description provided of the target population for which the PROM was developed?	Target population clearly described			Target population not clearly described	
4 Is a clear description provided of the context of use	Context of use clearly described		Context of use not clearly described		

<p>5 Was the PROM development study performed in a sample representing the target population for which the PROM was developed?</p>	<p>Study performed in a sample representing the target population</p>	<p>Assumable that the study was performed in a sample representing the target population, but not clearly described</p>	<p>Doubtful whether the study was performed in a sample representing the target population</p>	<p>Study not performed in a sample representing the target population (SKIP items 6-12)</p>	<p></p>
--	---	---	--	--	---------

<i>Concept elicitation (relevance and comprehensiveness)</i>		very good	adequate	doubtful	inadequate	NA
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?	Widely recognized or well justified qualitative method used, suitable for the construct and study population	Assumable that the qualitative method was appropriate and suitable for the construct and study population, but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was suitable for the construct and population	Method used not appropriate or not suitable for the construct or study population	
7	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience		Not applicable
8	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide	Not clear if a topic guide was used or doubtful if topic or interview guide		Not applicable

		was appropriate, but not clearly described	was appropriate or no guide		
--	--	---	--------------------------------	--	--

9	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings of interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
10	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	Not applicable
11	Was at least part of the data coded independently?	At least 50% of the data was coded by at least two researchers independently	11-49% of the data was coded by at least two researchers independently	Doubtful if two researchers were involved in the coding or only 1-10% of the data was coded by at least two researchers	Only one researcher was involved in coding or no coding	Not applicable

		independently				
12	Was data collection continued until saturation was reached?	Evidence provided that saturation was reached	Assumable that saturation was reached	Doubtful whether saturation was reached	Evidence suggests that saturation was not reached	Not applicable
13	For quantitative studies (surveys): was the sample size appropriate?	≥100	50-99	30-49	<30	Not applicable

1b. Cognitive interview study or other pilot test

		very good	adequate	doubtful	inadequate	NA
14	Was a cognitive interview study or other pilot test conducted?	YES			NO (SKIP items 15-35)	
<i>General design requirements</i>						
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?	Study performed in a sample representing the target population	Assumable that the study was performed in a sample representing the target population, but not clearly described	Doubtful whether the study was performed in a sample representing the target population	Study not performed in a sample representing the target population	
<i>Comprehensibility</i>						
16	Were patients asked about the comprehensibility of the PROM?	YES		Not clear (SKIP standards 17-25)	No (SKIP standards 17-	

<p>17 Were all items tested in their final form?</p>	<p>All items were tested in their final form</p>	<p>Assumable that all items were tested in their final form, but not clearly described</p>	<p>Not clear if all items were tested in their final form</p>	<p>25) Items were not tested in their final form or items were not re-tested after substantial adjustments</p>	
--	--	--	---	---	--

18	Was an appropriate qualitative method used to assess the comprehensibility of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was appropriate or not clear if patients were asked about the comprehensibility of the items, response options or recall period or patients not asked about the comprehensibility of the PROM instructions or the recall period	Method used not appropriate or patients not asked about the comprehensibility of the items or the response options	
19	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
20	Were skilled interviewers used?	Skilled group moderators/	Group moderators	Not clear if group moderators		Not applicable

<p>21 Were the interviews based on an appropriate interview guide?</p>	<p>interviewers used Appropriate topic or interview guide</p>	<p>/interviewers had limited experience or were trained specifically for the study Assumable that the topic or interview guide was appropriate, but not clearly described</p>	<p>/interviewers were trained or group moderators /interviewers not trained and no experience Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide</p>	<p></p>	<p>Not applicable</p>
<p>22 Were the interviews recorded and transcribed verbatim?</p>	<p>All group meetings or interviews were recorded and transcribed verbatim</p>	<p>Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described</p>	<p>Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews</p>	<p>No recording and no notes</p>	<p>Not applicable</p>

<p>23 Was an appropriate approach used to analyse the data?</p>	<p>A widely recognized or well justified approach was used</p>	<p>Assumable that the approach was appropriate, but not clearly described</p>	<p>Not clear what approach was used or doubtful whether the approach was appropriate</p>	<p>Approach not appropriate</p>	
<p>24 Were at least two researchers involved in the analysis?</p>	<p>At least two researchers involved in the analysis</p>	<p>Assumable that at least two researchers were involved in the analysis, but not clearly described</p>	<p>Not clear if two researchers were included in the analysis or only one researcher involved in the analysis</p>		
<p>25 Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?</p>	<p>No problems found or problems appropriately addressed and PROM was adapted and re-tested if necessary</p>	<p>Assumable that there were no problems or that problems were appropriately addressed, but not clearly described</p>	<p>Not clear if there were problems or doubtful if problems were appropriately addressed</p>	<p>Problems not appropriately addressed or PROM was adapted but items were not re-tested after substantial adjustments.</p>	<p>Not applicable</p>

<i>Comprehensiveness</i>		very good	adequate	doubtful	inadequate	NA
26	Were patients asked about the comprehensiveness of the PROM? _____	YES		NO or not clear (SKIP items 27-35)		
27	Was the final set of items tested?	The final set of items was tested	Assumable that the final set of items was tested, but not clearly described	Not clear if the final set of items was tested or not the final set of items was tested or the set of items was not re-tested after items were removed or added		
28	Was an appropriate method used for assessing the comprehensiveness of the PROM? _____	Widely recognized or well justified method used	Assumable that the method was appropriate but not clearly described or only quantitative (survey) method(s) used	Doubtful whether the method was appropriate or method used not appropriate		
29	Was each item tested in an appropriate number of patients? For qualitative studies	≥7	4-6	<4 or not clear		

For quantitative (survey) studies		≥50	≥30	<30 or not clear	
30	Were skilled interviewers used?	Skilled interviewers used	Interviewers had limited experience or were trained specifically for the study	Not clear if interviewers were trained or interviewers not trained and no experience	Not applicable

<p>31 Were the interviews based on an appropriate interview guide?</p>	<p>Appropriate topic or interview guide</p>	<p>Assumable that the topic or interview guide was appropriate, but not clearly described</p>	<p>Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide</p>		<p>Not applicable</p>
<p>32 Were the interviews recorded and transcribed verbatim?</p>	<p>All group meetings or interviews were recorded and transcribed verbatim</p>	<p>Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described</p>	<p>Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews or no recording and no notes</p>		<p>Not applicable</p>
<p>33 Was an appropriate approach used to analyse the data?</p>	<p>A widely recognized or well justified approach was used</p>	<p>Assumable that the approach was appropriate,</p>	<p>Not clear what approach was used or doubtful whether the approach was</p>		

<p>34 Were at least two researchers involved in the analysis?</p>	<p>but not clearly described</p> <p>appropriate or approach not appropriate</p> <p>At least two researchers involved in the analysis</p> <p>Assumable that at least two researchers were involved in the analysis, but not clearly described</p> <p>Not clear if two researchers were included in the analysis or only one researcher involved in the analysis</p>	
---	--	--

<p>35 Were problems regarding the comprehensiveness of the PROM appropriately addressed by adapting the PROM?</p>	<p>No problems found or problems appropriately addressed and PROM was adapted and re- tested if necessary</p>	<p>Assumable that there were no problems or that problems were appropriately addressed, but not clearly described</p>	<p>Not clear if there were problems or doubtful if problems were appropriately addressed or PROM was adapted but items were not re- tested after substantial adjustments</p>	<p>Problems not appropriately addressed</p>	<p>Not applicable</p>
---	---	---	--	---	---------------------------

Box 2. Content validity

2a. Asking patients about relevance

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
1	Was an appropriate method used to ask patients whether each item is relevant for their experience with the condition? _____	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if patients were asked whether each item is relevant or doubtful whether the method was appropriate	Method used not appropriate or patients not asked about the relevance of all items	
2	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
3	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators/interviewers had limited experience or were trained	Not clear if group moderators/interviewers were trained or group		Not applicable

<p>4 Were the group meetings or interviews based on an appropriate topic or interview guide?</p>	<p>Appropriate topic or interview guide</p>	<p>specifically for the study Assumable that the topic or interview guide was appropriate, but not clearly described</p>	<p>moderators /interviewers not trained and no experience Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide</p>		<p>Not applicable</p>
--	---	--	---	--	-----------------------

5	Were the group meetings or interviews recorded and transcribed verbatim?	All group Meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
<i>Analyses</i>						
6	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
7	Were at least two researchers involved in the analysis?	At least two researchers involved in the	Assumable that at least two researchers were	Not clear if two researchers were included in the		

	analysis	involved in the analysis, but not clearly described	analysis or only one researcher involved in the analysis	
--	----------	---	--	--

2b Asking patients about comprehensiveness

Design requirements

		very good	adequate	doubtful	inadequate	NA
8	Was an appropriate method used for assessing the comprehensiveness of the PROM? <hr/>	Widely Recognized or well justified method used	Only quantitative (survey) method(s) 99raine assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
9	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
10	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were 99raine dor group moderators /interviewers not		Not applicable

<p>11 Were the group meetings or interviews based on an appropriate topic or interview guide?</p>	<p>Appropriate Topic or interview guide</p>	<p>Assumable that the topic or interview guide was appropriate, but not clearly described</p>	<p>trained and no experience Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide</p>	<p>Not applicable</p>
---	---	---	--	-----------------------

12	Were the group meetings or interviews recorded and transcribed verbatim?	All group Meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
<i>Analyses</i>						
13	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
14	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the	Not clear if two researchers were included in the analysis or only one		

		analysis, but not clearly described	researcher involved in the analysis	
--	--	--	--	--

2c Asking patients about comprehensibility

Design requirements

		very good	adequate	doubtful	inadequate	NA
15	Was an appropriate qualitative method used for assessing the comprehensibility of the PROM instructions, items, response options, and recall period?	Widely recognized or	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) were used. It is doubtful whether the method was appropriate or not clear if patients were asked about the comprehensibility of the items, response options or recall period or patients not asked about the comprehensibility of the PROM Instructions	Method used not appropriate or patients not asked about the comprehensibility of the items, response options, or recall period	
16	Was each item tested in an appropriate number of patients?					

	For qualitative studies	≥7	4-6	<4 or not clear	
	For quantitative (survey) studies	≥50	≥30	<30 or not clear	
17	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators/interviewers had limited experience or were trained specifically for the study	Not clear if group moderators/interviewers were trained for group moderators/interviewers not trained and no experience	

<p>18 Were the group meetings or interviews based on an appropriate topic or interview guide?</p>	<p>Appropriate topic or interview guide</p>	<p>Assumable that the topic or interview guide was appropriate, but not clearly described</p>	<p>Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide</p>	<p></p>	<p>Not applicable</p>
<p>19 Were the group meetings or interviews recorded and transcribed verbatim?</p>	<p>All group meetings or interviews were recorded and transcribed verbatim</p>	<p>Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described</p>	<p>Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews</p>	<p>No recording and no notes</p>	<p>Not applicable</p>
<p><i>Analyses</i></p>					
<p>20 Was an appropriate approach used to analyse the data?</p>	<p>A widely recognized or</p>	<p>Assumable that the approach was</p>	<p>Not clear what approach was used</p>	<p>Approach not appropriate</p>	<p></p>

<p>21 Were at least two researchers involved in the analysis?</p>	<p>well justified approach was used</p>	<p>appropriate, but not clearly described</p>	<p>or doubtful whether the approach was appropriate</p>		
	<p>At least two researchers involved in the analysis</p>	<p>Assumable that at least two researchers were involved in the analysis, but not clearly described</p>	<p>Not clear if two researchers were included in the analysis or only one researcher involved in the analysis</p>		

2d. Asking professionals about relevance

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
22	Was an appropriate method used to ask professionals whether each item is relevant for the construct of interest? _____	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was	Not clear if professionals were asked whether each item is relevant or doubtful whether the method was appropriate	Method used not appropriate or professionals not asked about the relevance of all items	
23	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
24	Was each item tested in an appropriate number of professionals?					

	For qualitative studies	≥7	4-6	<4 or not clear	
	For quantitative (survey) studies	≥50	≥30	<30 or not clear	
<i>Analyses</i>					
25	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate

<p>26 Were at least two researchers involved in the analysis?</p>	<p>At least two researchers involved in the analysis</p>	<p>Assumable that at least two researchers were involved in the analysis, but not clearly described</p>	<p>Not clear if two researchers were included in the analysis or only one researcher involved in the analysis</p>	<p></p>
---	--	---	---	---------

2e. Asking professionals about comprehensiveness

Design requirement

		very good	adequate	doubtful	inadequate	NA
27	Was an appropriate method used for assessing the comprehensiveness of the PROM? <hr/>	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
28	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
29	Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		

<i>Analyses</i>						
30	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	

<p>31 Were at least two researchers involved in the analysis?</p>	<p>At least two researchers involved in the analysis</p>	<p>Assumable that at least two researchers were involved in the analysis, but not clearly described</p>	<p>Not clear if two researchers were included in the analysis or only one researcher involved in the analysis</p>	<p></p>
---	--	---	---	---------

Box 3. Structural validity

Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no

Does the study concern unidimensionality or structural validity? ² unidimensionality / structural validity

Statistical methods

		very good	adequate	doubtful	inadequate	NA
		Confirmatory factor				
1	For CTT: Was exploratory or confirmatory factor analysis performed?	analysis performed	Exploratory factor analysis performed		No exploratory or confirmatory factor analysis performed	Not applicable
2	For IRT/Rasch: does the chosen model fit to the research question?	Chosen model fits well to the research question	Assumable that the chosen model fits well to the research question	Doubtful if the chosen model fits well to the research question	Chosen model does not fit to the research question	Not applicable
3	Was the sample size included in the analysis adequate?	FA: 7 times the number of items and ≥ 100	FA: at least 5 times the number of items and ≥ 100 ; OR at least	FA: 5 times the number of items but < 100	FA: < 5 times the number of items	

		6 times number of items but <100		
Rasch/1PL models: ≥ 200 subjects	Rasch/1PL models: 100-199 subjects	Rasch/1PL models: 50-99 subjects	Rasch/1PL models: < 50 subjects	
2PL parametric IRT models OR Mokken scale analysis: ≥ 1000 subjects	2PL parametric IRT models OR Mokken scale analysis: 500- 999 subjects	2PL parametric IRT models OR Mokken scale analysis: 250- 499 subjects	2PL parametric IRT models OR Mokken scale analysis: < 250 subjects	

<i>Other</i>						
4 Were there any other important flaws in the design or statistical methods of the study?	<table border="0"> <tr> <td>No other important methodological flaws</td> <td style="background-color: #cccccc;">[Redacted]</td> <td>Other minor methodological flaws (e.g. rotation method not described)</td> <td>Other important methodological flaws (e.g. inappropriate rotation method)</td> <td style="background-color: #cccccc;">[Redacted]</td> </tr> </table>	No other important methodological flaws	[Redacted]	Other minor methodological flaws (e.g. rotation method not described)	Other important methodological flaws (e.g. inappropriate rotation method)	[Redacted]
No other important methodological flaws	[Redacted]	Other minor methodological flaws (e.g. rotation method not described)	Other important methodological flaws (e.g. inappropriate rotation method)	[Redacted]		

¹If the scale is not based on a reflective model, unidimensionality or structural validity is not relevant.

²In a systematic review, it is helpful to make a distinction between studies where factor analysis is performed on each (Desapriya *et al.*) scale separately to evaluate whether the (Desapriya *et al.*) scales are unidimensional (unidimensionality studies) and studies where factor analysis is performed on all items of an instrument to evaluate the (expected) number of subscales in the instrument and the clustering of items within subscales (structural validity studies).

Box 4. Internal consistency

Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no

<i>Design requirements</i>	very good	adequate	doubtful	inadequate	NA
1 Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistic calculated for each unidimensional scale or subscale		Unclear whether scale or sub scale is unidimensional	Internal consistency statistic NOT calculated for each unidimensional scale or sub scale	
<i>Statistical methods</i>					
2 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	Not applicable
3 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	Not applicable
4 For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation)	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	Not applicable

calculated?			
<i>Other</i>			
5 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws	Other minor methodological flaws	Other important methodological flaws

¹ If the scale is not based on a reflective model, internal consistency is not relevant

Box 5. Cross-cultural validity\Measurement invariance

<i>Design requirements</i>	very good	adequate	doubtful	inadequate	NA
1 Were the samples similar for relevant characteristics except for the group variable?	Evidence provided that samples were similar for relevant characteristics except group variable	Stated (but no evidence provided) that samples were similar for relevant characteristics except group variable	Unclear whether samples were similar for relevant characteristics except group variable	Samples were NOT similar for relevant characteristics except group variable	
<i>Statistical methods</i>					
2 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	Not applicable
3 Was the sample size included in the analysis adequate?	Regression analyses or IRT/Rasch based analyses: 200 subjects per group	150 subjects per group	100 subjects per group	< 100 subjects per group	

<p><i>Other</i></p> <p>4 Were there any other important flaws in the design or statistical methods of the study?</p>	<p>MGCFAs*: 7 times the number of items and ≥ 100</p> <p>No other important methodological flaws</p>	<p>5 times the number of items and ≥ 100; OR 5-7 times the number of items but < 100</p>	<p>5 times the number of items but < 100</p> <p>Other minor methodological flaws</p>	<p>< 5 times the number of items</p> <p>Other important methodological flaws</p>	
--	--	--	--	--	--

*MGCFAs: multi-group confirmatory factor analyses

Box 6. Reliability

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
1	Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
<i>Statistical methods</i>						
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal.	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided	No ICC or Pearson or Spearman correlations	Not applicable

5	For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated	<p>Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred</p> <p>[Redacted]</p>	<p>that no systematic change has occurred or WITH evidence that systematic change has occurred</p> <p>[Redacted]</p>	<p>calculated</p> <p>No kappa calculated</p>	<p>Not applicable</p>
---	---	------------------	---	--	--	-----------------------

6	For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated	[Redacted]	Unweighted Kappa calculated or not described	[Redacted]	Not applicable
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described	[Redacted]	[Redacted]	Not applicable
<i>Other</i>						
8	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws	[Redacted]	Other minor methodological flaws	Other important methodological flaws	[Redacted]

Box 7. Measurement error

<i>Design requirements</i>		very good	adequate	doubtful	Inadequate	NA
1	Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3	Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
<i>Statistical methods</i>						
4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (Al-Ani <i>et al.</i>) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's	Not applicable

5	For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		alpha, or on SD from another population % agreement not calculated	Not applicable
<i>Other</i>						
6	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 8. Criterion validity

		very good	adequate	doubtful	inadequate	NA
<i>Statistical methods</i>						
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
<i>Other</i>						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 9. Hypotheses testing for construct validity

9a. Comparison with other outcome measurement instruments (convergent validity)

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
1	Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
2	Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s), OR evidence for insufficient measurement properties of the comparator	

<i>Statistical methods</i>					instrument(s)	
		Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
3	Was the statistical method appropriate for the hypotheses to be tested?					

<i>Other</i>						
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws	

9b. Comparison between subgroups (discriminative or known-groups validity)

Design requirements

5 Was an adequate description provided of important characteristics of the subgroups?

	very good	adequate	doubtful	inadequate	NA
5	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
<i>Statistical methods</i>					
6	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
<i>Other</i>					
7	No other important methodological flaws		Other minor methodological flaws (e.g. only data presented on a comparison with an	Other important methodological flaws	



Box 10. Responsiveness

10a. Criterion approach (i.e. comparison to a gold standard)

		very good	adequate	doubtful	inadequate	NA
<i>Statistical methods</i>						
1	For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
<i>Other</i>						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
4	Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
5	Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s) OR evidence of poor quality of comparator instrument(s)	
<i>Statistical methods</i>						
6	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method were appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
<i>Other</i>						

<p>7 Were there any other important flaws in the design or statistical methods of the study?</p>	<p>No other important methodological flaws</p>		<p>Other minor methodological flaws</p>	<p>Other important methodological flaws</p>	
--	--	---	---	---	---

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
8	Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
<i>Statistical methods</i>						
9	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
<i>Other</i>						
10	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10d. Construct approach: (i.e. hypotheses testing: before and after intervention)

<i>Design requirements</i>		very good	adequate	doubtful	inadequate	NA
11	Was an adequate description provided of the intervention given?	Adequate description of the intervention		Poor description of the intervention	NO description of the intervention	
<i>Statistical methods</i>						
12	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
<i>Other</i>						
13	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

ANEXO II

Critérios para boas propriedades de medição

Propriedade de medida	Classificação ¹	Critérios
Validade estrutural	+	<p>CTT: CFA: CFI ou TLI ou medida comparável > 0,95 OU RMSEA < 0,06 OU SRMR < 0,08².</p> <p>IRT / Rasch: Nenhuma violação da <u>unidimensionalidade</u>³: CFI ou TLI ou medida comparável > 0,95 OU RMSEA < 0,06 OU SRMR < 0,08 E sem violação da <u>independência local</u>: correlações residuais entre os itens após o controle do fator dominante < 0,20 OU Q3 < 0,37 E sem violação da <u>monotonicidade</u>: gráficos com aparência adequada OU item escalabilidade > 0,30 E ajuste adequado <u>do modelo</u>: IRT: $\chi^2 > 0,01$ Rasch: quadrados médios de infit e equipamento $\geq 0,5$ e $\leq 1,5$ OR Z padronizados valores > -2 e < 2</p>
	?	<p>CTT: nem todas as informações de '+' foram relatadas IRT / Rasch: ajuste do modelo não relatado</p>
	-	Critérios para "+" não cumpridos
Consistência interna	+	Pelo menos evidência baixa ⁴ de validade estrutural suficiente ⁵ AND Alfa (α) de Cronbach $\geq 0,70$ para cada escala unidimensional ou subescala ⁶
	?	Critérios para "Pelo menos evidências baixas ⁴ para validade estrutural suficiente ⁵ não atendidas
	-	Pelo menos evidências baixas ⁴ para validade estrutural suficiente ⁵ E alfa (α)

		de <i>Cronbach</i> <0,70 para cada escala unidimensional ou subescala ⁶
Confiabilidade	+	ICC ou Kappa ponderado $\geq 0,70$
	?	ICC ou Kappa ponderado não relatado
	-	ICC ou Kappa ponderado <0,70
Erro de medida	+	SDC ou LoA <MIC ⁵
	?	MIC não definido
	-	SDC ou LoA > MIC ⁵
Teste de hipótese para validade de construto	+	O resultado está de acordo com a hipótese ⁷
	?	Nenhuma hipótese definida (pela equipe de revisão)
	-	O resultado não está de acordo com a hipótese ⁷
Validade transcultural \ invariância da medida	+	Não foram encontradas diferenças importantes entre os fatores do grupo (como idade, gênero, idioma) na análise de múltiplos fatores do grupo OU nenhuma DIF importante para os fatores do grupo (McFadden's $R^2 < 0,02$)
	?	Nenhuma análise fatorial de grupo múltiplo OU DIF realizada
	-	Diferenças importantes entre os fatores do grupo OU DIF foram encontrado
Validade de critério	+	Correlação com padrão-ouro $\geq 0,70$ OU AUC $\geq 0,70$
	?	Nem todas as informações para "+" foram relatadas
	-	Correlação com padrão-ouro <0,70 OU AUC <0,70
Responsividade	+	O resultado está de acordo com a hipótese ⁷ OU AUC $\geq 0,70$
	?	Nenhuma hipótese definida (pela equipe de revisão)
	-	O resultado não está de acordo com a hipótese ⁷ OU AUC < 0,70

AUC = Área Sob a cCurva, CFA = Análise Fatorial Confirmatória, CFI = Índice de Ajuste Comparativo, CTT = Teoria Clássica de Teste, DIF = Funcionamento Diferencial do Item, ICC = Coeficiente de Correlação Intraclasse, TRI = Teoria da Resposta ao Item, LoA = Limites de Concordância, MIC = Alteração Mínima Importante, RMSEA: Erro Médio Quadrático de Aproximação da Raiz, SEM = Erro Padrão de Medição, SDC = Menor Alteração Detectável, SRMR: Residuais Médios Raiz Padronizados, TLI = Índice Tucker-Lewis

¹ "+" = suficiente, "-" = insuficiente, "?" = Indeterminado.

² Para classificar a qualidade da pontuação resumida, as estruturas fatoriais devem ser iguais entre os estudos.

³ A unidimensionalidade refere-se a uma análise fatorial por subescala, enquanto a validade estrutural refere-se a uma análise fatorial de uma medida de resultado relatada pelo paciente (multidimensional).

⁴ Conforme definido pela classificação das evidências de acordo com a abordagem GRADE.

⁵ Essa evidência pode vir de diferentes estudos.

⁶ O critério "alfa de *Cronbach* < 0,95" foi excluído, pois é relevante na fase de desenvolvimento de uma PROM e não ao avaliar uma PROM existente.

⁷ Os resultados de todos os estudos devem ser considerados em conjunto e, em seguida, deve ser decidido se 75% dos resultados estão de acordo com as hipóteses.

ANEXO III

Abordagem GRADE

Qualidade da evidência	Rebaixamento
Alta	Risco de viés -1 Sério
Moderada	-2 Muito sério
Baixa	-3 Extremamente sério
Muito baixa	Inconsistência -1 Sério -2 Muito sério Imprecisão -1 total n = 50-100 -2 total n <50 Indirectividade -1 Sério -2 Muito sério

n= tamanho da amostra

(1) O risco de viés pode ocorrer se a qualidade do estudo for duvidosa ou inadequada, como avaliados com a lista de verificação de Risco de Viés do COSMIN ou se apenas um estudo de qualidade está disponível. A qualidade da evidência deve ser rebaixada com um nível (por exemplo, de evidência alta a moderada) se houver risco sério de viés, com duas (por exemplo, de alto a baixo), se houver risco muito sério de viés ou com três níveis (isto é, de alto a muito baixo) de risco extremamente elevado de viés. Na Tabela abaixo está exposto o que é considerado um risco grave, muito grave ou extremamente grave de viés.

Risco de viés	Rebaixamento para risco de viés
Não	Existem vários estudos de qualidade pelo menos adequada ou existe um estudo de muito boa qualidade disponível
Sério	Existem vários estudos de qualidade duvidosa disponíveis, ou existe apenas um estudo de qualidade adequada
Muito sério	Existem vários estudos de qualidade inadequada ou existe apenas um estudo de qualidade duvidosa disponível
Extremamente sério	Existe apenas um estudo de qualidade inadequada disponível

(2) Inconsistência: a inconsistência já pode ter sido resolvida através do agrupamento ou resumir os resultados de subgrupos de estudos com resultados semelhantes e fornecer classificações gerais para esses subgrupos. Nesse caso, não é necessário fazer o downgrade. Se nenhuma explicação para inconsistência for encontrada, a equipe de revisão poderá decidir não agrupar ou resumir resultados e classificá-los como "inconsistentes". Nesse caso, não será fornecido o nível de qualidade da evidência. No entanto, uma solução alternativa pode ser avaliar o resultado agrupado ou resumido (por exemplo, com base na maioria dos resultados) suficientes ou insuficientes e rebaixar a qualidade da evidência para inconsistência com um ou dois níveis. Os revisores também devem decidir o que será considerado grave (1 nível) ou muito grave (2 níveis) na inconsistência, porque isso depende do contexto. Cabe à equipe de revisão decidir qual parece ser a melhor solução em cada situação.

(3) Imprecisão refere-se à amostra total incluída nos estudos. Recomendamos que rebaixamento com um nível quando o tamanho total da amostra da estudos resumidos é inferior a 100 e com dois níveis quando o tamanho total da amostra abaixo de 50. Observe que este princípio não deve ser usado para medição de propriedades nas quais um requisito de tamanho de amostra já está incluído no COSMIN como Caixa de risco de viés, ou seja, validade de conteúdo, validade estrutural e validade transcultural.

(4) Indirectividade pode ocorrer se forem incluídos na revisão estudos que foram (parcialmente) realizados em outra população ou outro contexto de uso que não a

população ou o contexto de uso de interesse na revisão sistemática. Por exemplo, se apenas parte da população do estudo consistir em pacientes com a doença de interesse, a equipe de revisão poderá decidir fazer o rebaixamento com um ou dois níveis, para uma indireta séria ou muito séria. Pode-se considerar a desclassificação por indireta para validade de construto ou capacidade de resposta quando a evidência é considerada fraca. Por exemplo, quando a evidência é baseada apenas em comparações com PROMs que medem diferentes construções ou apenas em diferenças entre grupos extremamente diferentes. Como decidir sobre o que deve ser considerado sério ou muito sério na indirectividade depende do contexto e deve ser decidida pela equipe de revisão.