



UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE MEDICINA DA BAHIA
Mestrado em Saúde, Ambiente e Trabalho
Largo do Terreiro de Jesus – Centro Histórico,
40.026-010, Salvador, Bahia Brasil
Telefax: 55-71-3283-5572; 3283-5573; Cel. 8726-4059
e-mail: sat@ufba.br, ppgsat4@gmail.com
<http://www.sat.ufba.br>



VALIDADE E CONFIABILIDADE DO QUESTIONÁRIO NÓRDICO DE SINTOMAS MUSCULOESQUELÉTICOS: UMA REVISÃO SISTEMÁTICA DE LITERATURA

Ivone Batista Alves

Dissertação de Mestrado

Salvador (Bahia), 2017



UNIVERSIDADE FEDERAL DA BAHIA
FACULDADE DE MEDICINA DA BAHIA
Mestrado em Saúde, Ambiente e Trabalho
Largo do Terreiro de Jesus – Centro Histórico,
40.026-010, Salvador, Bahia Brasil
Telefax: 55-71-3283-5572; 3283-5573; Cel. 8726-4059
e-mail: sat@ufba.br, pLgsat4@gmail.com
<http://www.sat.ufba.br>



VALIDADE E CONFIABILIDADE DO QUESTIONÁRIO NÓRDICO DE SINTOMAS MUSCULOESQUELÉTICOS: UMA REVISÃO SISTEMÁTICA DE LITERATURA

Ivone Batista Alves

Professora-orientadora: Verônica Maria Cadena Lima

Dissertação apresentada ao Colegiado do Curso de Pós-graduação em Saúde, Ambiente e Trabalho da Faculdade de Medicina da Universidade Federal da Bahia, como pré-requisito para a obtenção do grau de Mestre em Saúde, Ambiente e Trabalho.

Salvador (Bahia), 2017

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),
com os dados fornecidos pelo(a) autor(a).

Batista Alves, Ivone

Validade e confiabilidade do questionário nórdico
de sintomas musculoesqueléticos: uma revisão
sistemática de literatura / Ivone Batista Alves. --
Salvador, 2017.

116 f. : il

Orientadora: Verônica Maria Cadena Lima.

Dissertação (Mestrado - Pós-graduação em Saúde,
Ambiente e Trabalho) -- Universidade Federal da
Bahia, Faculdade de Medicina, 2017.

1. Revisão sistemática. 2. Questionário Nórdico de
Sintomas Musculoesqueléticos. 3. Propriedades de
medida. 4. Validade e Confiabilidade.. I. Cadena
Lima, Verônica Maria. II. Título.

COMISSÃO EXAMINADORA

Membros Titulares:

Professora Dr^a Rita de Cássia Pereira Fernandes, professora adjunta do Departamento de Medicina Preventiva e Social da Faculdade de Medicina da Bahia da Universidade Federal da Bahia, doutora em Saúde Pública pelo Instituto de Saúde Coletiva da Universidade Federal da Bahia (2004), com estágio no exterior, bolsista CAPES, na *McGill University*, Montreal-Canadá.

Professora Dr^a Rita de Cássia Ribeiro Silva, professora associada do Departamento de Nutrição da Escola de Nutrição da Universidade Federal da Bahia. Professora do programa de pós-graduação em Alimentos, Nutrição e Saúde da Escola de Nutrição e do Programa de Pós-graduação em Saúde Coletiva do Instituto de Saúde Coletiva da UFBA, doutora em Saúde Coletiva pela UFBA (2001), pesquisadora CNPq - PROPI 2012.

Professora Dr^a Verônica Maria Cadena Lima (professora-orientadora), professora associada do Departamento de Estatística do Instituto de Matemática e Estatística da Universidade Federal da Bahia, professora do Programa de Pós-Graduação em Saúde, Ambiente e Trabalho e doutora em Estatística pela University of Leeds, Inglaterra (2003).

AGRADECIMENTOS

Ao longo desta caminhada, muitas pessoas foram essenciais para a conclusão desta importante etapa.

Agradeço primeiramente a Deus por permitir que chegasse até aqui.

À minha orientadora, Professora Verônica Lima, pela confiança, pela paciência, por todo o aprendizado e enriquecimento profissional e credibilidade depositadas em mim.

À professora Rita Rego, pelo apoio, pelo aprendizado e por ter sido a maior incentivadora para que chegasse até aqui.

À minha família pelo incentivo e confiança de sempre.

A Heider Macedo pelo companheirismo e apoio nos momentos difíceis.

As professoras da banca examinadora, Dr^a Rita Fernandes e Dr^a Rita Ribeiro pelas valiosas contribuições.

À Louise Machado pela amizade e por nunca me deixar desanimar.

Aos professores do Programa de Pós-Graduação em Saúde, Ambiente e Trabalho da Universidade Federal da Bahia pelos ensinamentos, à funcionária “Inha” e à ex-funcionária Solange Xavier pela amizade e apoio.

À secretária do PPGSAT, Caroline Casaes, pelo apoio na secretaria.

À equipe do grupo de pesquisa coordenado pela Professora Rita Rêgo.

Às colegas da Turma do Mestrado 2015, em especial, Patrícia Sandes, e Jamile Baultar, não somente pelas ajudas acadêmicas, mas pelo suporte e amizade.

SUMÁRIO

I. Resumo	8
II. Abstract	9
III. Objetivos	10
IV. Apresentação	11
V. Revisão de Literatura	13
V.1. Distúrbios Musculoesqueléticos Relacionados ao Trabalho	13
V.2. Questionário Nórdico de Sintomas Musculoesqueléticos	15
V.3. Validade e confiabilidade de um instrumento	18
VI. Artigo - Validade e Confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos: uma Revisão Sistemática de Literatura.	25
VI.1. Resumo	26
VI.2. Abstract	27
VI.3. Introdução	28
VI.4. Metodologia	30
VI.5. Resultados	36
VI.6. Discussão e Conclusão	48
VI.8. Referências Bibliográficas	54
VI.9. Tabelas	57
VII. Lista geral de referências bibliográficas	76
VIII. Anexos	80
VIII. 1. Anexo A- Versão Ampliada do Questionário de Sintomas Musculoesqueléticos.	80
VIII. 2. Anexo B - Checklist COSMIN	81
VIII. 3. Anexo C - Checklist COSMIN escala de 4 pontos.	90
VIII. 4. Anexo D - Protocolo da revisão	112
IX. Apêndice	116
IX. 1. Apêndice A- Estratégia de busca	

ÍNDICE DE TABELAS E FIGURAS

ARTIGO Validade e Confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos:
uma Revisão Sistemática de Literatura

TABELAS

Tabela 1.	CrITÉrios para avaliaÇo da qualidade das propriedades de medida.	57
Tabela 2.	CrITÉrio de avaliaÇo da qualidade metodolgica dos estudos de acordo com a escala de classificaÇo de quatro pontos por itens do checklist COSMIN.	58
Tabela 3.	CrITÉrios para determinar a qualidade geral das propriedades de medida de um instrumento.	61
Tabela 4.	Características dos estudos de validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos	62
Tabela 5.	Metodologia utilizada e regio anatmica avaliada nos estudos selecionados	67
Tabela 6.	AvaliaÇo da qualidade de medida validade de crITÉrio do Questionário Nórdico de Sintomas Musculoesqueléticos por regies anatmicas.	73
Tabela 7.	AvaliaÇo da qualidade de medida confiabilidade teste-reteste do Questionário Nórdico de Sintomas Musculoesqueléticos por regies anatmicas	74
Tabela 8.	AvaliaÇo da qualidade metodolgica dos estudos de validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) através do <i>checklist COSMIN</i>	75

FIGURA

Figura 1.	Fluxograma dos estudos selecionados.	37
------------------	--------------------------------------	----

I. RESUMO

O Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM), publicado em 1987, tem como finalidade servir como instrumento de vigilância dos Distúrbios Musculoesqueléticos em um contexto de ergonomia e para triagem na saúde ocupacional. É um instrumento frequentemente utilizado em estudos epidemiológicos para determinar a magnitude dos Distúrbios Musculoesqueléticos em diversos grupos ocupacionais no Brasil e no mundo. Esta dissertação intitulada “Validade e Confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM): Uma Revisão Sistemática de Literatura” tem como objetivo avaliar as propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM e a qualidade metodológica desses estudos. Os resultados mostraram que a literatura carece de estudos de boa qualidade metodológica referente à validade de critério e confiabilidade teste-reteste do QNSM. No que se refere ao planejamento amostral, ficou constatada a falta de clareza no cálculo do tamanho da amostra e a seleção não aleatória dos indivíduos das pesquisas. Diante do exposto, observa-se ausência de padronização metodológica, principalmente no planejamento amostral, para realização dos estudos de propriedades de medidas do QNSM. Sendo assim, espera-se que esta dissertação contribua com novos estudos de validade e confiabilidade de um instrumento de pesquisa.

Palavras-chave: Revisão sistemática, Questionário Nórdico de Sintomas Musculoesqueléticos, propriedades de medida, validade, confiabilidade.

II. ABSTRACT

The Nordic Questionnaire on Musculoskeletal Symptoms (QNSM), published in 1987, aims to serve as an instrument for monitoring Musculoskeletal Disorders in a context of ergonomics and for screening in occupational health. It is an instrument frequently used in epidemiological studies to determine the magnitude of Musculoskeletal Disorders in several occupational groups in Brazil and worldwide. This dissertation titled "Validity and Reliability of the Nordic Questionnaire of Musculoskeletal Symptoms (QMSM): A Systematic Review of Literature" aims to evaluate the measurement properties criterion validity and reliability of the NMQ and the methodological quality of these studies. The results showed that the literature lacks studies of good methodological quality regarding the criterion validity and reliability of the NMQ. Regarding the sampling plan used in these studies, it was not clear how the sample size was calculated and also there was not a random selection of the individuals to participate in the surveys. In conclusion, there is a lack of methodological standardization, especially in the sampling plans used to carry out evaluation of measurement properties of the NMQ. Therefore, we expect that this dissertation can contribute with new studies of the validity and reliability of a research instrument.

Key words: Systematic review, Nordic Musculoskeletal Symptom Questionnaire, measurement properties, validity, reliability.

III. OBJETIVOS

PRINCIPAL

Analisar os artigos publicados que avaliaram a validade e a confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM).

SECUNDÁRIOS

1. Identificar as regiões anatômicas analisadas nos artigos que avaliaram a validade de critério e confiabilidade teste-reteste do QNSM;
2. Avaliar as propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM;
3. Investigar a qualidade metodológica dos artigos selecionados; e,
4. Determinar a qualidade geral das propriedades validade de critério e confiabilidade teste-reteste do QNSM, por região anatômica.

IV. APRESENTAÇÃO

A temática desta dissertação surgiu da necessidade do grupo de pesquisa da qual participo de realizar um estudo sobre validade de critério do QNSM na população de marisqueiras de Saubara/BA para avaliar este questionário enquanto instrumento de triagem. Qual o tamanho da amostra necessário? Como calcular o tamanho da amostra? Qual padrão-ouro utilizar, já que não há padrão-ouro para estabelecer dor? Qual medida deve ser utilizada para quantificar a validade? Estas foram algumas das perguntas que precisavam ser respondidas para um adequado planejamento e desenvolvimento de um estudo deste tipo. Surgiu então a necessidade de realizar uma revisão sistemática de literatura para examinar estudos que avaliassem a validade de critério do QNSM em populações de trabalhadores. Com o objetivo de auxiliar em novos estudos, a confiabilidade teste-reteste do QNSM foi também examinada. Com isso, o objetivo desta dissertação é examinar os artigos publicados sobre validade de critério e confiabilidade teste-reteste do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) em população de trabalhadores. Em particular, nesta dissertação, consideramos as propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM, por região anatômica. Com o conhecimento adquirido nesta revisão sistemática, esperamos estar aptos para realizar estudos futuros sobre validade e confiabilidade de um instrumento de medida.

Após selecionar os artigos, inicialmente as características dos estudos foram sintetizadas: autores; ano; local; propriedades de medida avaliadas; grupo estudado; padrão-ouro; e versão do QNSM utilizada. Posteriormente, as propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM por região anatômica foram

classificadas para cada um dos artigos. Foi realizada uma avaliação sobre a qualidade metodológica destes artigos através do Checklist COSMIN (*Consensus-based Standards for the Selection of Health Measurement Instruments*) e uma classificação geral do QNSM foi obtida por região anatômica.

Nesta dissertação, discute-se a metodologia utilizada nos estudos de validade de critério e confiabilidade teste-reteste do QNSM e a importância da utilização de métodos estatísticos adequados para a realização de pesquisa científica tais como descrição de critérios para o cálculo do tamanho da amostra e seleção aleatória dos indivíduos.

V. REVISÃO DE LITERATURA

V.1 DISTÚRBIOS MUSCULOESQUELÉTICOS RELACIONADOS AO TRABALHO

Estudos realizados em muitos países vêm mostrando que os Distúrbios Musculoesqueléticos (DME) relacionados ao trabalho são um importante problema de saúde pública, apresentando alta prevalência em países industrializados e em desenvolvimento (FERNANDES et al., 2016; FERNANDES et al., 2011; MARRAS et al., 2009). Os DME além de gerarem incapacidade e sofrimento para o trabalhador, resultam em altos custos para a sociedade (MARRAS et al., 2009). Eles são definidos como lesões ou disfunções que afetam os músculos, tendões, ligamentos, articulações, nervos periféricos e raízes nervosas, em diferentes segmentos corporais como a região do pescoço, ombro, cotovelos, antebraços, punhos, mãos, região lombar e segmentos dos membros inferiores (COSTA e VIEIRA, 2010; PUNNETT e WEGMAN, 2004).

Os riscos para desenvolvimento dos DME são multifatoriais (MARRAS et al., 2009; NRC e IM, 2001), e incluem fatores individuais (aumento do índice de massa corporal, idade), fatores físicos relacionados ao trabalho (movimentos repetitivos, trabalho pesado, flexão frequente, vibração do corpo inteiro, posturas inadequadas) e fatores psicossociais (alta demanda e baixo controle sobre o trabalho) (COSTA e VIEIRA 2010; PUNNETT e WEGMAN, 2004).

A região lombar e membros superiores são as regiões do corpo mais sujeitas a riscos para desenvolvimento de DME relacionados ao trabalho (MARRAS et al., 2009). Estudos realizados pelo *National Research Council – NRC, National Institute for*

Occupational Safety and Health - NIOSH e o *National Institute of Health - NIH* dos EUA relataram que a dor na região lombar é uma importante causa de morbidade e incapacidade no mundo (NRC e IM, 2001).

Na Europa, os DME são os principais problemas que causam incapacidade para o trabalho, principalmente aqueles relacionados à dor nas costas e membros superiores, gerando custo de até 2% do produto interno bruto (ZHELTOUKHOVA, 2013).

Nos Estados Unidos, dados do *Bureau of Labor Statistics*, mostram que em 2015 do total de casos de DME em trabalhadores, 31% representaram casos de entorses ou estirpe resultante de levantamento de pesos. Os DME exigiram um tempo mediano de 12 dias de ausência ao trabalho para recuperação dos trabalhadores da indústria privada americana em todo ano de 2015.

No Brasil, o Instituto Nacional de Seguro Social (INSS), em 2016, registrou 15.624 benefícios por incapacidade laboral causada por DME nas costas (Dorsalgias) e 13.187 por lesões no ombro, representando um quantitativo de 7% e 6% respectivamente dos benefícios por incapacidade concedida pelo órgão no referido ano (<http://www.previdencia.gov.br/dados-abertos/estatsticas/tabelas-cid-10/>). Os dados do INSS são subnotificados, pois não contemplam os trabalhadores afastados por incapacidade por menos de 15 dias e ainda como relata Fernandes e colaboradores (2011) esses dados são restritos aos trabalhadores do emprego formal.

O sistema de vigilância em DME no ambiente ocupacional é direcionado para identificação, prevenção e redução dos seus fatores de risco (KUORINKA e FORCIER, 1995).

O autorrelato é um procedimento mais informativo em relação ao exame físico na identificação dos DME (PUNNET e WEGMAN, 2004).

V.2 QUESTIONÁRIO NÓRDICO DE SINTOMAS MUSCULOESQUELÉTICOS (QNSM)

O Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) foi desenvolvido por um projeto apoiado pelo Conselho de Ministros Nórdicos (órgão interparlamentar criado em 1971), envolvendo os países Nórdicos (Dinamarca, Finlândia, Islândia, Noruega e Suécia). Este questionário teve como objetivos principais: servir como instrumento de vigilância dos distúrbios musculoesqueléticos em um contexto de ergonomia; e, para triagem nos serviços de saúde ocupacional (KUORINKA et al.,1987). O QNSM foi padronizado para facilitar a comparação dos resultados de diferentes estudos sobre análise e registro de sintomas musculoesqueléticos relacionados ao trabalho (KUORINKA et al.,1987; PINHEIRO et al.,2002), sendo constituído de duas partes: uma geral e outra específica. A parte geral compreende um levantamento simples, com objetivo de avaliar a presença de dor ou desconforto nos últimos 12 meses em nove regiões anatômicas do corpo (pescoço, ombros, cotovelos, punho/mão, parte alta das costas, região lombar, quadris/coxas, nádegas/joelhos, tornozelos/ pés) (KUORINKA et al., 1987). A parte específica do questionário, que teve inicialmente foco na região lombar, pescoço e ombro, faz uma avaliação mais profunda analisando a gravidade dos sintomas em termos de consequências sobre as atividades no trabalho e no lazer, e duração total dos sintomas, absenteísmo ao trabalho nos últimos 12 meses e 7 dias anterior a entrevista (KUORINKA et al., 1987).

A literatura aponta algumas limitações técnicas para o uso do QNSM tais como, a experiência do entrevistador e tempo dos sintomas de DME. Segundo Kuorinka e colaboradores (1987), a experiência do entrevistador pode afetar os resultados, sintomas mais recentes e graves tendem a ser lembrados mais facilmente do que os mais velhos e menos graves. Para Anderson e colaboradores (1997) a forma como o questionário é aplicado influencia na confiabilidade de uma pesquisa sobre sintomas. Em um estudo com motoristas de ônibus com objetivo de avaliar os efeitos da forma de aplicação da versão preliminar do QNSM, os autores encontraram maior prevalência de sintomas de DME quando o questionário foi aplicado no contexto de avaliação no ambiente de trabalho em relação ao contexto de saúde geral dos trabalhadores (FERNANDES et al., 2011; BARON et al., 1996; ANDERSON et al., 1987). Quando aplicado de forma padronizada para todos os trabalhadores e se o objetivo da investigação for para determinar áreas prioritárias para programas de intervenção, a forma na qual o instrumento é preenchido não invalida os resultados da pesquisa (BARON et al., 1996).

Como a sensibilidade das pesquisas sobre DME normalmente é alta e sua especificidade é baixa isso pode afetar a precisão das respostas (FERNANDES et al., 2011). Com o objetivo de minimizar o erro na definição de casos e melhorar a especificidade do instrumento, o QNSM foi ampliado com inclusão de questões sobre severidade, duração e frequências dos sintomas em segmento corporal pré estabelecido (KUORINKA E FORCIER, 1995; BARON et al., 1996; FERNANDES et al., 2011).

O QNSM foi traduzido para diversos idiomas (PINHEIRO et al., 2002) e tem sido amplamente utilizado em estudos epidemiológicos para identificar a prevalência de DME em diversas regiões do corpo, para várias categorias profissionais e em diferentes países, tais como: Dinamarca (PAARUP et al., 2011), EUA (LEMASTERS et al.,

1998), França (PAROT-SCHINKE et al., 2012), Grécia (ALEXOPOULOS et al., 2006), Irã (GHASEMKHANI, 2008), Reino Unido (HOUSSAIAN et al., 2004), Tailândia (JAWANTANAKUL et al., 2008) e Brasil (FALCÃO et al., 2015; PATARO e FERNANDES, 2014).

Avaliar a validade e confiabilidade de instrumentos utilizados para fins epidemiológicos pode fornecer informações úteis para subsidiar no levantamento dos sintomas de DME em programas de triagem na saúde ocupacional e contribuir na prevenção através de tratamento precoce (BARON et al., 1996). Para garantir a boa qualidade dos dados e permitir precisão na estimativa da prevalência de sintomas entre populações é importante utilizar instrumento epidemiológico validado (FERNANDES et al., 2011).

Estudos sobre a prevalência de DME em trabalhadores que utilizaram o QNSM encontraram altas prevalências. Em relação à dor musculoesquelética em uma ou mais regiões do corpo em pescadores artesanais/marisqueiras, a prevalência foi de 94,7% (FALCÃO et al, 2015) e em trabalhadores da limpeza urbana 62,8% (PATARO E FERNANDES, 2014).

A definição de “casos de DME” a partir do QNSM avalia dor ou desconforto em nove regiões anatômicas do corpo, nos últimos doze meses de trabalho, com duração mínima de uma semana ou frequência mínima mensal, não causada por lesão aguda. Sintomas associados a, pelo menos, um dos seguintes itens de gravidade: grau de severidade maior ou igual a três, em uma escala de 0 a 5 (nenhum desconforto a dor insuportável); busca de atenção médica pelo problema; ausência ao trabalho (oficial ou não); ou mudança de trabalho por restrição de saúde (KUORINKA & FORCIER, 1995; FERNANDES et al, 2010).

V.3 VALIDADE E CONFIABILIDADE DE UM INSTRUMENTO

Validade e confiabilidade são atributos importantes, que garantem a qualidade de um instrumento de medida (MUKKINK et al., 2009, SCHOLTES, 2011). Um instrumento para ser válido tem que ser confiável (FULLERTON, 1993). Se a validade do mesmo é questionável, aumenta o risco de conclusões errôneas (FULLERTON, 1993) contribuindo para graves consequências na tomada de decisões.

Validade refere-se ao grau e a precisão em que um instrumento mede o que se propõe a medir (MONTEIRO, 2014; SCHOLTES, 2011; FULLERTON, 1993). A validade de um instrumento pode ser avaliada de três formas, são elas: validade de conteúdo, validade de construto e validade em relação a um critério.

A validade de conteúdo tem como objetivo verificar se os itens de um instrumento refletem adequadamente determinado conteúdo a ser medido (SCHOLTES et al., 2011, 2014; FULLERTON, 1993). O conteúdo medido deve seguir os seguintes aspectos relacionados com o desenvolvimento do instrumento: o objetivo da medição deve ser claro (discriminativo avaliativo ou assistido); os conceitos os quais ele foi concebido para medir; os itens devem ser julgados e medidos por vários especialistas de acordo com sua relevância e abrangência; o conteúdo do instrumento deve ser específico para a população-alvo da pesquisa (TERWEE et al., 2007; SCHOLTES et al., 2011). Desta forma, a validade de conteúdo está relacionada com o processo de construção de um instrumento, em que verifica se os itens do instrumento estão medindo de fato o que se pretende medir, identificando se há discrepâncias e omissões no instrumento, avaliando a compreensão e clareza dos itens.

A validação de conteúdo não é expressa por escore ou por coeficiente quantitativo e sim por lógica, de forma subjetiva, através de um painel de especialistas qualificados no domínio de interesse (MONTEIRO, 2014; SCHOLTES et al., 2011; FULLERTON, 1993). Terwee e colaboradores (2007) relatam que a validade de conteúdo é considerada insuficiente se não descrever claramente o objetivo da medição, a população-alvo, os conceitos que estão sendo medidos e a seleção dos itens. Para Fullerton (1993) uma forma de avaliar quantitativamente a validade de conteúdo é verificar a concordância entre avaliadores.

A validade de construto determina o grau em que os escores de um instrumento de medição são consistentes com hipóteses baseadas na suposição de que ele mede o construto de interesse (SCHOLTES et al., 2011). Ela inclui validade estrutural e validade transcultural (SCHOLTES et al., 2011). A validade estrutural estima o grau em que os escores de um instrumento refletem adequadamente a dimensão do construto medido. Validade transcultural estima o grau em que o resultado do desempenho dos itens do instrumento traduzido ou adaptado culturalmente reflete adequadamente os resultados da versão original do instrumento (SCHOLTES et al., 2011).

Segundo Fullerton (1993), um teste que possui validade de construto deve ser altamente associado com outro de mesma característica (validação convergente) e ter pouca associação com outro de característica diferente (validação discriminante). A validade estrutural pode ser determinada pela técnica de análise fatorial e não há consenso sobre o número de hipóteses que devem ser testadas e/ou confirmadas para garantir uma validação adequada (TERWEE et al., 2007; SCHOLTES et al., 2011). Segundo Terwee e colaboradores (2007) para se ter uma validação de construto

adequada o ideal é que no mínimo 75% dos resultados devem corresponder às hipóteses que precisam ser, obrigatoriamente, especificadas com antecedência em um grupo de pelo menos 50 sujeitos.

A validação em relação a um critério avalia se o resultado de um instrumento concorda com o resultado de outro considerado como padrão-ouro (MONTEIRO, 2014; FULLERTON, 1993). A validade em relação ao um critério se refere à validade concorrente e a validade preditiva. Na validade concorrente a aplicação de um novo instrumento e do padrão-ouro é feita simultaneamente para verificar a correlação entre eles, em um curto intervalo de tempo (MONTEIRO, 2014; FULLERTON, 1993). Enquanto na validade preditiva o critério estará disponível no futuro para predizer ou generalizar o desempenho do instrumento (MONTEIRO, 2014; FULLERTON, 1993).

Para Fullerton (1993), a validade concorrente tem aplicação importante na tomada de decisões, este tipo de validade só faz sentido se o novo instrumento for vantajoso em relação ao padrão-ouro, como ser mais seguro, mais preciso, mais confortável, mais barato e possível de ser aplicado numa fase precoce da doença (MONTEIRO, 2014). A validade preditiva de um instrumento é assegurada, quando o mesmo pode ser usado como uma ferramenta de rastreio na seleção e classificação de doentes para programas de intervenção clínica (FULLERTON, 1993). A estimação da validade de critério depende do tipo de variável medida pelo instrumento e do padrão-ouro (SCHOLTES et al., 2011). No caso em que ambos são de escala contínua utiliza-se o coeficiente de correlação; se o instrumento tem escala contínua e o padrão-ouro possui escala dicotômica, a área sobre a curva ROC (*Receiver Operating Characteristic*) deverá ser avaliada; e, se ambas as escalas de medida forem dicotômicas, utiliza-se a

sensibilidade e a especificidade (SCHOLTES et al., 2011). No caso em que as medidas têm escalas contínuas, alguns autores apontam que para um critério ser considerado válido, a correlação entre o instrumento e o padrão-ouro deve ser de pelo menos 0,70 (TERWEE et al., 2007; SCHOLTES et al., 2011).

Os indicadores de validade mais utilizados para dados dicotômicos são sensibilidade, especificidade e valores preditivos. A sensibilidade de um teste refere-se à proporção de doentes com resultado do teste positivo, enquanto que a especificidade refere-se à proporção de não doentes com resultado do teste negativo (MURPHY, 1993). O valor preditivo positivo refere-se à probabilidade de pacientes com resultados do teste positivo, os verdadeiros doentes, terem a doença, dentre todos que tiveram o resultado do teste positivo (MURPHY, 1993). Já o valor preditivo negativo refere-se à probabilidade de pacientes com resultados do teste negativo, os verdadeiros não doentes, não terem a doença dentre todos com resultados do teste negativo (MURPHY, 1993). A sensibilidade e especificidade são atributos do instrumento, sua utilização é útil para avaliar validade preditiva (MONTEIRO, 2014; BARON et al., 1996). Um programa de rastreamento que pretende detectar os verdadeiros doentes o ideal é que se tenha um instrumento altamente sensível, enquanto quando se pretende uma confirmação diagnóstica o ideal é um instrumento altamente específico (MONTEIRO, 2014).

Confiabilidade é definida como o grau em que um instrumento fornece medidas repetidas com resultados iguais, ou seja, é quanto um instrumento está livre de erros (TERWEE et al., 2007; SCHOLTES et al., 2011). A confiabilidade pode ser estimada das seguintes formas: consistência interna; teste-reteste (entre avaliadores e intra-avaliadores) e erros de medição (SCHOLTES et al., 2011).

A consistência interna é uma forma de avaliar a confiabilidade de um instrumento quando não há repetições de medidas (SCHOLTES et al., 2011). Ela está relacionada com a homogeneidade das medidas. Neste caso, as pontuações dos itens de um instrumento são resumidas em uma pontuação total, sendo importante quando um instrumento pretende medir o mesmo conceito através de vários itens (TERWEE et al., 2007; SCHOLTES et al., 2011). Antes de estimar a consistência interna é necessário verificar, através da análise fatorial, a hipótese de que os itens de um instrumento medem um único conceito (SCHOLTES et al., 2011). Uma medida adequada de consistência interna é o alfa de Cronbach (TERWEE et al., 2007) que quantifica a correlação entre os itens da escala. Um valor do alfa entre 0,70 e 0,95 é um bom indicador de confiabilidade. O alfa de Cronbach é calculado depois de ser determinado o número de subescalas homogêneas pela análise fatorial, é calculado para cada subescala separadamente (TERWEE et al., 2007) e pode ser utilizado por variáveis dicotômicas, categóricas ou contínuas (MONTEIRO, 2014).

O método teste-reteste utiliza repetições de medidas para avaliar a confiabilidade do instrumento e pode ser estimada de duas maneiras, entre-avaliadores e dentre-avaliadores (SCHOLTES et al., 2011). Neste caso, a confiabilidade está relacionada com a reprodutibilidade das medidas. Na confiabilidade entre-avaliadores, as medidas de um mesmo instrumento são aferidas por dois avaliadores ao mesmo tempo, para o mesmo paciente e verifica-se se elas são concordantes entre os avaliadores (SCHOLTES et al., 2011). Enquanto que na confiabilidade dentre-avaliadores, as medidas de um instrumento são aferidas por um único avaliador em intervalos distintos de tempo e verifica-se se a medida sofreu mudança ao decorrer do tempo (SCHOLTES et al., 2011). Para Scholtes e colaboradores (2011), o intervalo de tempo ideal para

aferir as medidas deve ser o suficiente para que a última pontuação não afete a primeira, ou seja, por conta do viés de memória, ele deve ser típico de duas a quatro semanas. Intervalo curto superestima a confiabilidade enquanto intervalo longo subestima.

Segundo Scholtes e colaboradores (2011), as estimativas de confiabilidade teste-reteste dependem do tipo de dados, no caso de variáveis categóricas o percentual de concordância é feito utilizando o cálculo do kappa ponderado de Cohen. Já para dados contínuos, utiliza-se a correlação intra-classe (SCHOLTES et al., 2011). Para ambas as estimativas, o ideal é que sejam superiores a 0,70 (Terwee et al., 2007).

No processo de avaliação da confiabilidade de um instrumento, erros de medição associados às mudanças na pontuação de uma mesma medida de um instrumento, podem ocorrer ao ser aplicado repetidamente (FULLERTON, 1993). É importante que o pesquisador conheça os erros de medição associados ao instrumento de pesquisa e selecione o instrumento com o menor erro (FULLERTON, 1993). Segundo Fullerton e colaboradores (1993) as fontes de erros mais comuns são: erros devidos ao tempo de amostragem; erros associados com uso de itens específicos e erros associados à amostragem.

O erro devido ao tempo de amostragem é quando o pesquisador deseja utilizar o instrumento para fazer medição em ocasiões diferentes. Neste caso, a verdadeira pontuação pode ser afetada por mudança ao longo do tempo (FULLERTON, 1993). O erro associado com uso de itens específicos é aquele devido ao vocabulário, os itens selecionados para análise podem ser difíceis de serem compreendidos pelo entrevistado (FULLERTON, 1993). Já o erro associado à amostragem é aquele devido apenas ao processo da amostra. A repetição de medidas pode confirmar o resultado inicial ou

apontar para discrepâncias que podem ser devidas a erros aleatórios (FULLERTON, 1993).

VI. ARTIGO

VALIDADE E CONFIABILIDADE DO QUESTIONÁRIO NÓRDICO DE
SINTOMAS MUSCULOESQUELÉTICOS: UMA REVISÃO SISTEMÁTICA DE
LITERATURA

Ivone Batista Alves

VI. 1. RESUMO

Objetivo: Este artigo tem como objetivo avaliar os artigos publicados sobre validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) em populações de trabalhadores. **Metodologia:** Foram selecionados estudos nas bases de dados eletrônica *CINAHL*, *MEDLINE/PubMed*, *Scopus*, e *Web of Science* que satisfizeram os seguintes critérios de inclusão: Estudos de validação de critério e/ou confiabilidade teste-reteste do QNSM aplicados a população de trabalhadores; publicados em inglês ou português no período de 1987 a 2017. A avaliação da qualidade metodológica dos estudos selecionados foi realizada por dois examinadores independentes por meio do Checklist COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments). A qualidade das medidas relativas à validade de critério e confiabilidade teste-reteste foram avaliadas e a qualidade geral do QNSM, por região anatômica, foi determinada. **Resultados:** A estratégia de busca inicial identificou 194 estudos, sendo 15 artigos incluídos na revisão. De acordo com o Checklist COSMIN, nenhum estudo tanto de validade de critério quanto confiabilidade obteve qualidade metodológica classificada como boa ou excelente. A qualidade de medida validade de critério foi avaliada como sem classificação, enquanto a confiabilidade teste-reteste como indeterminada. A qualidade geral do QNSM foi classificada como conflitante para validade e indeterminada para confiabilidade. **Conclusão:** A validade e confiabilidade dos artigos sobre o QNSM não apresentam boa qualidade metodológica. Novos estudos sobre validade de critério e confiabilidade teste-reteste do QNSM devem ser realizadas em novas categorias ocupacionais.

Palavras-chave: Revisão sistemática, Questionário Nórdico de Sintomas Musculoesqueléticos, propriedades de medida, validade, confiabilidade.

VI.2. ABSTRACT

Objective: This article aims to evaluate the validity and reliability of the Nordic Musculoskeletal Symptom Questionnaire (QNSM) in worker populations. **Methods:** We selected studies in the electronic databases CINAHL, MEDLINE/PubMed, Scopus and Web of Science that met the following inclusion criteria: criterion validity and/or test-retest reliability studies of QNSM applied to worker population; published in English or Portuguese during 1987 to 2017. The evaluation of the methodological quality was performed by two independent examiners through the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) Checklist. The quality of measures regarding criterion validity and test-retest reliability were evaluated and the overall quality of the NMQ for each anatomic region were obtained. **Results:** The initial search strategy identified 194 studies, 15 of which were included in the review. According to the COSMIN Checklist, no validity or reliability study obtained methodological quality classified as good or excellent. The quality of the validity measure was evaluated as no information available while the test-retest reliability was classified as indeterminate rating. The overall quality for the QNSM was classified as conflicting for validity and unknown for reliability. **Conclusion:** The validity and reliability studies of the QNSM seem to lack of an adequate methodological quality. New studies of criterion validity and test-retest reliability of the QNSM should be performed for a variety of occupations.

Keywords: Systematic review, Nordic Musculoskeletal Symptom Questionnaire, measurement properties, validity, reliability.

VI.3. INTRODUÇÃO

Os distúrbios musculoesqueléticos relacionados ao trabalho (DME) são definidos como lesões ou disfunções que afetam os músculos, tendões, ligamentos, articulações, nervos periféricos e raízes nervosas, em diferentes segmentos corporais como a região do pescoço, ombro, cotovelos, antebraços, punhos, mãos, região lombar e segmentos dos membros inferiores (COSTA e VIEIRA, 2010; PUNNETT e WEGMAN, 2004).

Os riscos para desenvolvimento dos DME são decorrentes de vários fatores (MARRAS et al., 2009) e incluem fatores individuais (aumento do índice de massa corporal-IMC, idade), fatores físicos relacionados ao trabalho (movimentos repetitivos, trabalho pesado, flexão frequente, vibração do corpo inteiro, posturas inadequadas) e fatores psicossociais (alta demanda e baixo controle sobre o trabalho) (COSTA e VIEIRA 2010; PUNNETT e WEGMAN, 2004).

O Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) é uma ferramenta que tem sido bastante utilizada em levantamentos epidemiológicos (PALMER et al. 1999), para coletar informações sobre DME. Este questionário foi criado com o objetivo de servir como instrumento de triagem em um contexto de ergonomia e ser usado no serviço de saúde ocupacional, podendo ser auto-administrável ou aplicado através de entrevistas (KUORINKA et al.,1987).

O QNSM é constituído de duas partes: uma geral e outra específica. A parte geral compreende um levantamento simples, com objetivo de avaliar a presença de dor ou desconforto nos últimos 12 meses em nove regiões anatômicas do corpo (pescoço,

ombros, cotovelos, punho/mão, parte alta das costas, região lombar, quadris/coxas, nádegas/joelhos, tornozelos/ pés) (KUORINKA et al., 1987). Já a parte específica do QNSM, que teve inicialmente foco na região lombar, pescoço e ombro, faz uma avaliação mais profunda dos sintomas analisando sua gravidade em termos de consequências para atividades no trabalho e no lazer, e sua duração total, absenteísmo ao trabalho nos últimos 12 meses e 7 dias anterior a entrevista (KUORINKA et al., 1987; BARON et al., 1996). Este questionário foi ampliado com questões sobre severidade dos sintomas, medido em uma escala de 0 a 5, em que 0 indica sem dor e 5 dor severa; tempo de duração e frequência dos sintomas em segmento corporal pré-estabelecido (KUORINKA E FORCIER, 1995; BARON et al., 1996).

O QNSM é um instrumento bastante utilizado no Brasil e no mundo para estimar a prevalência de DME em diversos grupos ocupacionais. Ele foi padronizado para facilitar a comparação dos resultados de diferentes estudos sobre análise e registro de DME relacionados ao trabalho (KUORINKA et al., 1987; PINHEIRO et al., 2002).

Revisões sistemáticas têm sido realizadas para avaliar as propriedades de medida de diferentes instrumentos em populações específicas (TERWEE et al., 2007). Se esses estudos apresentam boa qualidade metodológica, suas conclusões são mais confiáveis (MOKKINK et al., 2010a). Para isso é necessário que alguns requisitos sejam atendidos, tais como tamanho da amostra adequado, utilização de padrão-ouro correto nos estudos de validade de critério, métodos estatísticos adequados para análise dos resultados e intervalo de tempo apropriado entre aplicações do instrumento nos estudos de confiabilidade teste-reteste (TERWEE et al., 2007).

Quando o instrumento é usado como ferramenta de rastreio para seleção e classificação é importante que ele possua uma validade de critério adequada (FULLERTON, 1993). O investigador tem que deixar evidente que o instrumento utilizado na pesquisa é válido e confiável (FULLERTON, 1993).

Pesquisas realizadas sobre validade e confiabilidade do QNSM em diversas categorias ocupacionais têm concluído que este questionário é uma ferramenta válida e confiável para ser utilizado na vigilância de saúde do trabalhador. Entretanto, não existem informações na literatura sobre o nível de qualidade das propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM e sobre a qualidade metodológica destes estudos. Desta forma, o objetivo deste estudo é realizar uma revisão sistemática de literatura sobre validade de critério e confiabilidade teste-reteste do QNSM de modo a avaliar a qualidade das propriedades de medida validade de critério e confiabilidade teste-reteste deste questionário, por região anatômica.

VI. 4. METODOLOGIA

Uma revisão sistemática de literatura foi realizada para identificar artigos publicados que avaliaram validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) em população de trabalhadores. O presente estudo foi registrado no *International Prospective Register of Systematic Reviews* – PROSPERO: nº CRD42016051213.

Os artigos foram selecionados nas bases de dados eletrônicas *CINAHL*, *MEDLINE/PubMed*, *Scopus*, e *Web of Science*. A busca foi realizada a partir de setembro de 1987 até janeiro de 2017, período após a publicação do QNSM. Em cada base foram utilizadas diferentes combinações dos termos de pesquisa (Apêndice A). A

Estratégia de busca incluiu o uso de vocabulário técnico-científico com termos *MeSH* (“*Medical Subject Headings*”) e DeCS (“*Descritores em Ciências da Saúde*”): *Validity OR Validation OR Sensibility OR Specificity OR Reproducibility OR Repeatability OR Reliability And Nordic Musculoskeletal Questionnaire*. Uma busca manual foi feita para selecionar estudos através das referências de artigos.

Após a identificação dos artigos através de seus títulos e resumos, aqueles que satisfizeram os seguintes critérios de inclusão foram selecionados: Estudos de validação e/ou confiabilidade do QNSM; e Artigos publicados em inglês ou português. Foram excluídos estudos sobre a versão preliminar do QNSM e revisão sistemática. Em seguida, os artigos foram lidos na íntegra por dois avaliadores e selecionados de forma independentes.

Os estudos selecionados tiveram os seguintes dados extraídos: autores, ano de publicação, país, título, grupo ocupacional avaliado, propriedades de medida investigadas, metodologia utilizada. Foram realizadas avaliações da qualidade das propriedades de medida relativas à validade de critério e confiabilidade teste-reteste segundo os critérios disponíveis por Terwee et al., (2007). A qualidade metodológica foi avaliada de acordo com *COSMIN checklist COnsensus-based Standards for the selection of health status Measurement INstruments* (MOKKINK et al., 2010a; TERWEE et al., 2012). Posteriormente, a qualidade geral do QNSM, por região anatômica, foi determinada de acordo com os critérios seguidos por Elbers e colaboradores (2012).

Avaliação da qualidade das propriedades de medida

A Tabela 1 descreve os critérios utilizados na avaliação da qualidade das propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM. Esses critérios foram sugeridos por Terwee et al., (2007). Em cada estudo, as propriedades avaliadas recebem uma classificação como adequada (+), não-adequada (-), indeterminada (?) ou sem classificação (0), dependendo do desenho de estudo, metodologia e resultados. Se não existe uma clara descrição de como o estudo foi planejado, a propriedade de medida avaliada é classificada como indeterminada. A validade de critério é classificada como adequada se existem argumentos que o padrão-ouro é “ouro”, ou seja, o padrão-ouro utilizado na pesquisa pode ser considerado como adequado e se a correlação do padrão-ouro com o instrumento é, no mínimo, 0,70. Os pesquisadores do COSMIN consideram inadequado a utilização de outro questionário como padrão-ouro, sendo a única exceção quando um instrumento, de tamanho menor, é comparado com a sua versão original longa, esta pode ser considerada como padrão-ouro. Apesar de não existir um padrão para dor, baseado no COSMIN, neste artigo foi considerado como padrão-ouro inadequado, a utilização dos resultados de outro questionário para comparar com os resultados do QNSM. Desta forma, qualquer outra maneira de diagnóstico foi considerada como um padrão-ouro adequado para avaliar a validade de critério.

A validade de critério inclui validade preditiva e validade concorrente. Na validade concorrente a aplicação de um novo instrumento e do padrão-ouro é realizada simultaneamente para verificar a correlação entre eles, em um curto intervalo de tempo. Na validade preditiva, o critério estará disponível no futuro para predizer ou generalizar o desempenho do instrumento (MONTEIRO, 2014; FULLERTON, 1993). Neste

estudo, a validade de critério foi avaliada considerando as informações disponibilizadas nos artigos sobre validade preditiva ou validade concorrente.

Na avaliação da confiabilidade teste-reteste, se a correlação intra-classe (ICC) ou o kappa for superior ou igual a 0,70, a qualidade desta medida será avaliada como adequada (Tabela 1). Segundo Terwee e colaboradores (2007), se o desenho do estudo não estiver claramente descrito ou se houver falhas importantes na sua metodologia ou execução (como por exemplo, viés de seleção ou uma população bastante heterogênea), as propriedades de medidas avaliadas são classificadas como indeterminadas.

Avaliação da qualidade metodológica

A avaliação da qualidade metodológica dos artigos selecionados foi realizada com base no COSMIN. Esta lista de verificação foi desenvolvida por um grupo internacional de especialistas, em várias rodadas de discussão, das áreas de epidemiologia, estatística, psicologia e clínica médica. (MOKKINK et al., 2010a; TERWEE et al., 2012). O objetivo inicial do *checklist* COSMIN foi avaliar a qualidade metodológica de estudos sobre o instrumento *Health Related Patient Reported Outcomes-HR-PROS (Resultados Relacionados à Saúde Relatados por Pacientes)*, mas o *checklist* COSMIN pode ser utilizado também para avaliar a qualidade metodológica de outros instrumentos de medida, pois as propriedades de medida avaliadas são as mesmas entre os instrumentos utilizados na área de saúde (MOKKINK et al., 2010a; TERWEE et al., 2012).

A taxonomia COSMIN é baseada em três domínios distintos, validade, confiabilidade e responsividade (MOKKINK et al., 2010a). A confiabilidade contém

três propriedades de medida: consistência interna, confiabilidade e erro de medida; a validade também possui três propriedades de medida: validade de conteúdo, validade de construto e validade de critério. A responsividade, que é a capacidade de um instrumento detectar mudanças clínicas importantes em certo período de tempo, contém apenas uma propriedade de medida, a qual é também chamada de responsividade (MOKKINK et al., 2010a). Neste artigo, vamos avaliar a qualidade metodológica dos estudos sobre as propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM baseado no checklist COSMIN.

O checklist COSMIN é formado por 12 seções, em que 10 são utilizadas para avaliar se o estudo atende às normas de boa qualidade metodológica (TERWEE et al., 2012; MOKKINK et al., 2010a) (ver Anexo B). Destas, nove são relativas às propriedades de medida: consistência interna (seção A), confiabilidade (seção B), erro de medida (seção C), validade de conteúdo (seção D), validade estrutural (seção E), teste de hipóteses (seção F), validade transcultural (seção G), validade de critério (seção H) e responsividade (seção I). Uma seção adicional foi incluída para avaliar a qualidade metodológica do estudo sobre interpretabilidade (seção J) (MOKKINK et al., 2010a; TERWEE et al., 2012). O checklist COSMIN possui ainda duas outras seções: uma para estudos que utilizam a metodologia de Teoria de Resposta ao Item (seção IRT) e outra para avaliar a generalização dos resultados de um estudo sobre propriedades de medida (seção de generalização) (TERWEE et al., 2012; MOKKINK et al., 2010a MOKKINK et al., 2010).

Para completar o checklist COSMIN são necessários quatro passos: 1. Identificar quais propriedades de medida estão sendo avaliadas em um artigo; 2. Determinar se o

método estatístico utilizado no artigo é baseado na Teoria dos Testes Clássicos ou na Teoria de Resposta ao Item. No caso do último método, a seção Teoria de Resposta ao Item deve ser preenchida. 3. Avaliar a qualidade metodológica dos estudos relativo as propriedades de medida identificadas no primeiro passo; e, 4. Preencher a seção referente aos requisitos sobre a generalização dos resultados, através da escala de quatro pontos que classifica cada propriedade de medida como excelente, boa, razoável ou pobre (MOKKINK et al., 2010a)

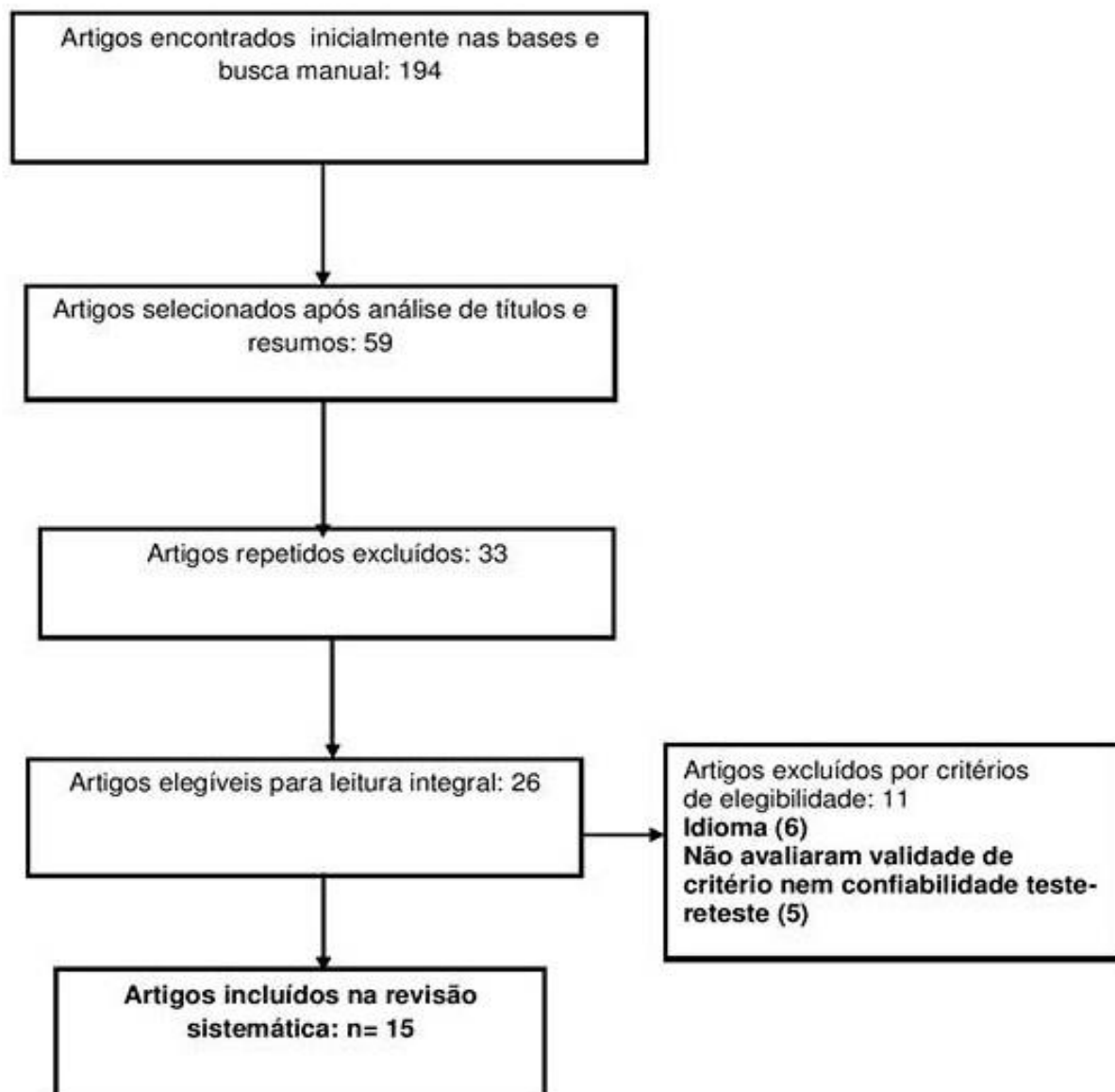
A Tabela 2 apresenta os critérios para avaliar a qualidade metodológica dos artigos, em relação às propriedades validade de critério e confiabilidade teste-reteste (TERWEE et al., 2012). Inicialmente, cada um dos sete itens da seção de validade de critério e os 14 itens referentes à seção de confiabilidade teste-reteste receberam as respostas “sim”, “não” ou “não aplicável”. Posteriormente, cada um dos itens foi avaliado recebendo a classificação: “excelente”, “boa”, “razoável” ou “pobre”, de acordo com os critérios presentes na Tabela 2. Uma classificação global da qualidade metodológica dos artigos para validade de critério e/ou confiabilidade teste-reteste foi obtida levando em consideração a classificação mais baixa obtida pelo estudo entre os itens avaliados em cada propriedade (TERWEE et al., 2012).

Após avaliar a qualidade das propriedades de medidas pelos critérios sugeridos por Terwee et al., (2007) e a qualidade metodológica dos estudos pelo *Checklist COSMIN*. Uma medida síntese foi utilizada para determinar a qualidade geral das medidas de validade de critério e confiabilidade teste-reteste do QNSM para cada região corpo, com base na Tabela 3 (ELBERS et al., 2012). Para esta medida síntese, Elbers e colaboradores (2012) propõem levar em consideração: a classificação obtida para

qualidade das propriedades de medida avaliada; a consistência dos resultados entre estudos; a qualidade metodológica dos estudos; e o número de estudos que investigaram a propriedade de medida (ELBERS et al., 2012). A qualidade geral das propriedades validade de critério e confiabilidade teste-reteste do QNSM, para cada região do corpo, foi classificada como forte, moderada, limitada, conflitante ou desconhecida conforme os critérios descritos na Tabela 2.

VI. 4.1. RESULTADOS

De acordo com a estratégia de busca utilizada nas bases e com a busca manual, foram identificados 194 artigos. Após aplicar os critérios de exigibilidade e excluir os artigos repetidos, 15 artigos relevantes foram selecionados para compor a revisão sistemática (Figura 1).

Figura 1: Fluxograma da seleção dos artigos

A Tabela 4 apresenta as características dos estudos selecionados referente a autores, ano de publicação, país, propriedades de medidas avaliadas, grupo estudado, definição de caso, definição do padrão-ouro e versão do QNSM utilizada na pesquisa. A publicação dos artigos selecionados compreendeu o período de 1994 a 2016. Nove artigos investigaram apenas validade (TAKEKAWA et al., 2015; PERREAULT et al., 2008; DESCATHA et al., 2007; PINHEIRO et al., 2002; AKESSON et al., 1999; BJORKSTEN et al., 1998; SILVERSTEIN et al., 1997; STAL et al., 1997; OHLSSON et al., 1994), quatro apenas confiabilidade (KAHRAMAN et al., 2016; NAMNIK et al., 2015; KNUDSEN et al., 2014; BARROS e ALEXANDRE, 2003) e dois avaliaram validade e confiabilidade (MESQUITA, et al., 2010; PALMER et al., 1999). Dentre os estudos de validade, dois avaliaram a validade concorrente (MESQUITA, et al., 2010; 2008; PINHEIRO et al., 2002) e nove a validade preditiva. O tamanho da amostra dos estudos variou de 39 a 2.685 indivíduos.

Os grupos ocupacionais avaliados foram trabalhadores da indústria, de um setor administrativo, de um armazém de alimentos, de uma indústria de automóvel; profissionais de um hospital; residentes de cirurgia ortopédica; bancários; profissionais de odontologia, professores universitários de um departamento de enfermagem e trabalhadores de ordenha.

Os estudos foram realizados no Brasil, Canadá, Estados Unidos, França, Irã, Portugal, Suécia e Turquia. O Brasil e a Suécia foram os países que apresentaram o maior número de estudos com 4 artigos cada (BARROS e ALEXANDRE, 2003; PALMER et al., 1999; PINHEIRO et al., 2002; TAKEKAWA et al., 2015) e (AKESSON et al., 1999; BJORKSTEN et al., 1998; STAL et al., 1997; OHLSSON et

al., 1994) respectivamente. Em seguida, EUA com 2 artigos (KNUDSEN et al., 2014; SILVERSTEIN et al., 1997) e Canadá (PERREAULT et al., 2008); França (DESCATHA et al., 2007); Irã (NAMNIK et al., 2015); Portugal (MESQUITA, et al., 2010); e Turquia (KAHRAMAN et al., 2016) tiveram um estudo cada.

No Brasil, Takekawa e colaboradores (2015) estudaram a validade de critério do QNSM, do *Roland-Morris Disability Questionnaire* (RDQ) e da Escala Visual Analógica (VAS), considerando uma variável contínua variando de 0 a 100 mm. A pesquisa foi realizada em uma amostra de 50 funcionários de um setor administrativo de uma universidade pública. Como padrão-ouro foi utilizado o resultado de uma avaliação física que incluiu teste de mobilidade, força, palpação e testes neurológicos para coluna lombar. Os autores avaliaram a sensibilidade e especificidade do QNSM comparando o resultado da avaliação física (padrão-ouro) com as respostas das seguintes perguntas: Presença de dor nos últimos 12 meses; Limitações funcionais para realizar trabalho, atividades domésticas e lazer devido à dor lombar; Procura por cuidados médicos devido à dor lombar e sintomas nos últimos 7 dias.

Barros e Alexandre (2003) em uma pesquisa de validação transcultural do QNSM, realizada no Brasil, avaliaram sua confiabilidade teste-reteste. Em uma amostra de 40 indivíduos distribuídos em 4 grupos distintos (alunos de graduação em enfermagem, professores do departamento de enfermagem, funcionários administrativos da universidade e pacientes atendidos pelo serviço de fisioterapia em um hospital universitário). Os autores utilizaram as perguntas do QNSM relativa à dor e desconforto nos últimos 12 meses e nos últimos sete dias nas nove regiões anatômicas (pescoço, ombro, parte superior das costas, cotovelos, punhos/mãos, quadril/coxa, joelhos,

tornozelos/pé); limitações para realização de tarefas por causa da dor e procura por atenção médica devido ao problema. A confiabilidade foi avaliada através do coeficiente kappa e o intervalo entre as aplicações do questionário foi de um dia.

Pinheiro e colaboradores (2002) avaliaram a validade concorrente do QNSM, em uma amostra de 90 bancários. A história clínica e o diagnóstico relatado nos prontuários dos trabalhadores, considerado como padrão-ouro, foram comparados com os relatos de sintomas nos últimos 12 meses e /ou 7 dias; 12 meses e/ou afastamentos das atividades rotineiras em nove regiões anatômicas (pescoço, ombro, cotovelos, punhos/mãos, parte altas das costas, região lombar, quadris/coxas, joelhos, tornozelos/pés) no QNSM. Os autores calcularam o coeficiente de correlação de Pearson entre os relatos de sintomas e/ou afastamento no QNSM e afastamento por sintomas na história clínica.

Palmer e colaboradores (1999) avaliaram a validade e a confiabilidade do QNSM em 105 pacientes ambulatoriais com o objetivo de discutir os resultados para a prática da saúde ocupacional. Os autores utilizaram o kappa de Cohen para avaliar a confiabilidade com intervalo de uma semana entre as aplicações do QNSM. A sensibilidade e especificidade foram avaliadas utilizando a pergunta sobre dor nos membros superiores, pescoço, joelhos, e quadris com duração de um dia ou mais na semana passada e no ano passado; dor na região citada impedindo realização de atividades diárias normais ao longo do ano passado. Como padrão-ouro foi utilizado o diagnóstico de um reumatologista.

Akesson e colaboradores (1999) em um estudo de seguimento na Suécia avaliaram a sensibilidade e a especificidade do QNSM em uma amostra de 90 profissionais de odontologia (30 dentistas, 30 assistentes e 30 higienistas dentários).

Neste estudo, os autores utilizaram a pergunta referente à dor e desconforto no pescoço ou ombro nos últimos 12 meses e 7 dias anterior a entrevista. Como padrão-ouro foi utilizado o diagnóstico de um exame físico padronizado.

Bjorksten e colaboradores (1998) avaliaram a validade do QNSM em uma amostra de 171 trabalhadoras de empresa de processamento de alimentos e serviço industrial da Suécia. Os autores avaliaram dor e desconforto no pescoço e ombro nos últimos 3 meses e 7 dias. O padrão-ouro utilizado foi baseado em critérios utilizados no diagnóstico de um fisioterapeuta através de uma avaliação clínica.

Stal e colaboradores (1997) realizaram um estudo de seguimento e avaliaram a sensibilidade e especificidade do QNSM em uma amostra de 161 trabalhadoras da ordenha na Suécia. Dor, dor e desconforto nos últimos 12 meses e 7 dias nos punhos/mãos foram à pergunta do QNSM utilizada no estudo. O padrão-ouro foi o diagnóstico clínico definido através de resultados de um exame físico padronizado.

Ohlsson e colaboradores (1994) avaliaram a validade do QNSM através da sensibilidade e especificidade numa amostra de 165 mulheres trabalhadoras de uma fábrica em ocupações variadas, na Suécia. Os autores avaliaram dor, dor ou desconforto no pescoço, ombro, cotovelo e mãos nos últimos 7 dias antecedentes a entrevista. O padrão-ouro utilizado foi o diagnóstico estabelecido através de critérios no exame físico.

Perreault e colaboradores (2008) realizaram um estudo no Canadá com 187 trabalhadores de escritório com o objetivo de avaliar a concordância entre o QNSM e o exame físico. Foi avaliada a presença de sintomas na região cervical (pescoço-ombro) por pelo menos 3 dias nos últimos 7 dias ou intensidade de dor superior a 50 mm em

uma Escala Visual Analógica de 100 mm. A sensibilidade e especificidade foram calculadas comparando diferentes definições de casos de sintomas do QNSM e do diagnóstico do exame físico realizado depois de 21 dias ou menos e mais de 21 dias.

Descatha e colaboradores (2007) na França, através de um estudo de validade compararam os resultados do QNSM em duas pesquisas sobre DME em membros superiores (ombro, braço, cotovelo, antebraço, punho e mão). A primeira pesquisa era baseada em 1.757 trabalhadores que preencheram o QNSM no período de 1993-1994 e 598 no período de 1996-1997. Na segunda pesquisa, foram entrevistados 2.685 trabalhadores no período de 2002-2003, que faziam parte de sistema de vigilância epidemiológica de DME em membros superiores (ombro, braço, cotovelo, antebraço, punho e mão). Os trabalhadores foram submetidos a exames clínicos padronizados e seus resultados foram comparados com a pergunta do QNSM relacionada à dor e desconforto nos membros superiores (ombro, braço, cotovelo, antebraço, punho e mão) nos últimos 6 meses anterior a entrevista, na primeira pesquisa. Na segunda pesquisa os trabalhadores reportaram a dor e desconforto nos membros superiores (ombro, braço, cotovelo, antebraço, punho e mão) nos últimos 12 meses. Os autores definiram como caso de DME a presença de pelo menos um sintoma no ombro, braço, cotovelo, antebraço ou punho e mão. Os autores, através do índice kappa, verificaram a concordância dos sintomas relatados pelo QNSM com os resultados do exame clínico. Uma escala variando de 0 a 10 foi utilizada para verificar intensidade da dor e através da curva *ROC (Receiver Operating Characteristic)* foi avaliado qual o melhor ponto de corte da sensibilidade e especificidade para discriminar corretamente os doentes.

Em estudo realizado no Irã, Namnik e colaboradores (2015) avaliaram a confiabilidade do QNSM em uma amostra de 122 trabalhadores industriais com relação aos DME nas regiões lombar e joelho. Os autores justificaram que por falta de um padrão para se definir DME, a validade foi relatada em termos de validade de constructo. A confiabilidade teste-reteste foi avaliada utilizando o coeficiente kappa com intervalo entre a aplicação do QNSM de uma semana.

Em Portugal, Mesquita e colaboradores (2010) realizaram um estudo de adaptação transcultural do QNSM e avaliaram confiabilidade teste-reteste e validade concorrente em uma amostra de 60 trabalhadores de um armazém de alimentos. Os autores utilizaram as perguntas: dor, dor e desconforto nas nove regiões anatômicas (pescoço, ombro, cotovelos, punhos/mãos, parte altas das costas, região lombar, quadris/coxas, joelhos, tornozelos/pés) nos últimos 12 meses e 7 dias; e impedimento de realizar tarefas diárias por causa do problema. A confiabilidade foi avaliada utilizando o coeficiente kappa num intervalo de 1 semana entre as aplicações do QNSM. A validade concorrente foi avaliada utilizando como padrão-ouro o índice de incapacidade *Oswestry*, segundo os autores, este é um questionário que mede o impacto de deficiência da dor lombar.

Na Turquia, Kahraman e colaboradores (2016) realizaram um estudo de adaptação transcultural e avaliaram a confiabilidade do QNSM em uma amostra de 198 profissionais de um hospital universitário (74 secretárias, 60 enfermeiras, 48 cuidadores e trabalhadores da limpeza e 18 médicos). A confiabilidade foi avaliada utilizando o coeficiente kappa com intervalo de 1 semana entre as aplicações do QNSM.

Knudsen e colaboradores (2014) realizaram um estudo de confiabilidade do QNSM em uma amostra de 39 cirurgiões ortopédicos residentes nos EUA. Os autores calcularam o kappa para avaliar a confiabilidade do QNSM, que foi aplicado entre um intervalo de quatro semanas. Neste estudo foram utilizadas perguntas sobre sintomas nos últimos 12 meses e 7 dias nas nove regiões anatômicas (pescoço, ombro, cotovelos, punhos/mãos, parte altas das costas, região lombar, quadris/coxas, joelhos, tornozelos/pés); e se houve impedimento para realização de atividades rotineira nos últimos 12 meses.

Silverstein e colaboradores (1997), em um estudo de validade com 626 trabalhadores, analisaram dados de quatro fábricas de automóveis em fases iniciais de implementação de programas de ergonomia. Os trabalhadores responderam ao QNSM e em seguida foram submetidos ao exame físico. Os autores utilizaram como caso de DME a presença de sintomas que tenham ocorrido pelo menos 20 vezes durante o último ano e que durasse, no mínimo, uma semana. Como padrão-ouro foi utilizado os resultados de um exame físico.

Em todos os artigos, os autores foram claros ao descrever as características dos indivíduos participantes da pesquisa no que se refere às informações sobre idade, distribuição por sexo, configuração em que a pesquisa foi realizada e objetivo do estudo.

As regiões anatômicas analisadas nos artigos selecionados estão apresentadas na Tabela 5. Nesta tabela apresentamos também as medidas de sensibilidade, especificidade, correlação tau-de-Kendal e coeficiente de correlação de Pearson, para as regiões anatômicas avaliadas. As regiões mais avaliadas foram ombro e pescoço, sendo

avaliados por 10 e 11 dos 15 estudos, respectivamente. Enquanto parte altas das costas, tornozelos/pés e quadril/coxa foram as regiões menos avaliadas nos artigos. Somente Descatha e colaboradores (2007) e Perreault e colaboradores (2008) analisaram as regiões considerando unidades funcionais, membros superiores (ombros, braços, cotovelos, antebraços, punho e mão) e região cervical (pescoço/ombro).

A Tabela 6 apresenta a avaliação da qualidade de medida validade de critério de acordo com que foi estabelecido na Tabela 1. A qualidade de medida validade de critério foi considerada como “sem classificação” para 9 artigos (TAKEKAWA et al., 2015; PERREAULT et al., 2008; DESCATHA et al., 2007; PALMER et al., (1999); AKESSON et al., 1999; BJORKSTEN et al., 1998; STAL et al., 1997; SILVERSTEIN et al., 1997; OHLSSON et al., 1994). Para estes artigos, a descrição da metodologia apresentada foi insuficiente uma vez que a avaliação da qualidade de medida da validade critério é limitada ao valor da correlação entre o padrão-ouro e o instrumento. Por exemplo, no estudo de Takekawa et al., (2015), os autores não utilizaram a correlação para avaliar validade de critério. Mesquita e colaboradores (2010) e Pinheiro e colaboradores (2002) tiveram a qualidade de medida indeterminada em relação à validade de critério. Nos estudos de Mesquita e colaboradores (2010) o que influenciou para tal classificação foi a utilização de padrão-ouro considerando inadequado, segundo os critérios utilizados na avaliação. Enquanto no estudo de Pinheiro e colaboradores (2002), apesar dos autores utilizarem um padrão-ouro considerado adequado, a qualidade de medida validade de critério foi considerada indeterminada devido à falta de clareza da descrição metodológica. Os atores não deixaram claro sobre a realização do cálculo do tamanho da amostra e seleção aleatória dos indivíduos.

Na Tabela 7 é apresentada a avaliação da qualidade de medida de confiabilidade teste-reteste do QNSM para as nove regiões do corpo usando os critérios de qualidade apresentados na Tabela 1. Do total de seis artigos, apenas um, o de Knudsen e colaboradores (2014), apresentou qualidade da medida de confiabilidade teste-reteste adequado. Nesse estudo, o QNSM foi classificado como adequado para avaliar distúrbios musculoesqueléticos nas regiões do pescoço, cotovelos, região lombar, parte alta das costas, punhos/mãos, quadril/coxa e joelhos. E não adequado para as regiões do ombro e tornozelos/pé porque os valores do índice de concordância kappa foram inferiores a 0,70 para essas regiões. Já maioria dos estudos (Kahraman et al., 2016; Namnik et al., 2015; Barros e Alexandre, 2003; Mesquita et al., 2010; e Palmer et al., 1999) utilizaram o intervalo de tempo entre as aplicações do QNSM inferior a duas semanas, quando o tempo considerado como ideal é de duas a quatro semanas. Isso deixou dúvidas quanto à metodologia utilizada nesses estudos. Por isso a classificação deles quanto à qualidade da confiabilidade teste-reteste foi especificada como indeterminada para as regiões anatômicas avaliadas (Tabela 6).

A Tabela 8 apresenta a avaliação da qualidade metodológica dos estudos em relação à validade de critério e confiabilidade teste-reteste através do *checklist COSMIN*. Os artigos apresentaram qualidade metodológica classificada como razoável ou pobre, tanto para validade de critério quanto para confiabilidade.

Dos onze estudos que avaliaram a validade de critério, dez (Pinheiro et al., 2002; Palmer et al., 1999; Takekawa et al., 2015; Bjorksten et al., 1998; Akesson et al., 1999; Stal et al., 1997; Ohlsson et al., 1994; Silverstein et al., 1997; Perreault et al., 2008; Descatha et al., 2007) apresentaram qualidade metodológica classificada como razoável.

O que impediu o estudo de Palmer e colaboradores (1999) de ser classificado como excelente e os outros estudos como bom foi o item: Houve alguma falha importante na concepção ou métodos do estudo? Todos os estudos de validade de critério apresentaram falha no seu método devido à falta de clareza sobre o cálculo da amostra e/ou seleção dos participantes da pesquisa.

Apenas o artigo de Mesquita e colaboradores (2010) foi classificado como pobre, porque além de não deixar claro quanto ao cálculo da amostra e da seleção dos participantes, não utilizou o padrão-ouro considerado como adequado. Isso é considerado uma falha importante na avaliação da qualidade metodológica nos estudos de validade de critério visto que outro instrumento não pode ser considerado como padrão-ouro (SCHOLTES et al., 2011).

Dos seis estudos que avaliaram a confiabilidade teste-reteste do QNSM, cinco (KAHRAMAN et al., 2016; NAMNIK et al., 2015; BARROS e ALEXANDRE, 2003; MESQUITA et al., 2010; PALMER et al., 1999) tiveram sua qualidade metodológica classificada como pobre porque não utilizaram um intervalo adequado entre as aplicações do questionário. Knudsen e colaboradores (2014) foi o único estudo analisado que utilizou o tempo adequado para reaplicar o QNSM. Sua qualidade metodológica foi classificada como razoável pelo fato de não mencionar como o tamanho da amostra foi obtido.

Por fim, os critérios estabelecidos por Elbers et al., (2011) e descritos na Tabela 2 foram utilizados para avaliar a qualidade geral das propriedades de medida relativas a validade de critério e confiabilidade teste-reteste do QNSM para todas as regiões anatômicas. A qualidade geral para a validade de critério do QNSM foi classificada

como conflitante, para todas as regiões anatômicas, pois um grande número de estudos apresentou qualidade de medida para a validade de critério como sem classificação ou qualidade metodológica razoável.

Em relação à confiabilidade do QNSM, a maioria dos estudos apresentou qualidade de medida indeterminada e qualidade metodológica pobre. Com isso, a qualidade geral para a confiabilidade teste-reteste do QNSM, para todas as regiões anatômicas, é classificada como desconhecida.

VI. 6. Discussão e Conclusão

Este estudo revisou sistematicamente 15 artigos sobre validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM). A qualidade das propriedades de medidas relativas à validade de critério e a confiabilidade teste-reteste do QNSM foram avaliadas de acordo com Terwee et al., (2007) e, por meio do *checklist* COSMIN, a qualidade metodológica destes estudos foi examinada. Uma classificação geral das propriedades de medida do QNSM foi obtida para cada região anatômica com base em Elbers et. al., (2012).

Onze estudos avaliaram a validade e seis a confiabilidade. A maioria dos artigos que avaliaram a validade de critério (PINHEIRO et al., 2002; PALMER et al.,1999; TAKEKAWA et al.,2015; Bjorksten et al.,1998; AKESSON et al.,1999; STAL et al., 1997; OHLSSON et al., 1994; SILVERSTEIN et al., 1997; PERREAULT et al.,2008; DESCATHA et al.,2007) apresentou qualidade metodológica classificada como razoável. Com relação à qualidade metodológica do estudo de Mesquita e colaboradores (2010) foi classificado como pobre porque não utilizou um padrão-ouro considerado como adequado.

Um ponto forte dos artigos que avaliaram a validade de critério foi o fato de que a maioria deles utilizarem um padrão-ouro adequado. Foram utilizados: avaliação física que incluiu teste de mobilidade, força, palpação e testes neurológicos para região lombar (Takekawa et al., 2015); diminuição $\geq 30\%$ da amplitude ativa de movimento ou diminuição da força muscular normal (pontuação ≤ 4 na escala de Lovett) ou dor com intensidade ≥ 3 na escala de 11 pontos no local relevante durante qualquer manobra (PERREAULT et al., 2008); diagnóstico realizado por um médico do trabalho através da análise da história clínica e diagnóstico do prontuário do trabalhador na empresa (PINHEIRO et al., 2002); exame clínico padronizado (DESCATHA et al., 2007; AKESSON et al., 1999; BJORKSTEN et al., 1998; STAL et al., 1997; OHLSSON et al., 1994); diagnóstico realizado por um reumatologista de espondilose cervical, capsulite adesiva e epicondilite lateral (PALMER et al., 1999); dor no local da queixa durante a manobra, agravada por testes de movimento resistido (SILVERSTEIN et al., 1997). Apenas Mesquita e colaboradores (2010) utilizaram um padrão-ouro incorreto. Os autores compararam os resultados do QNSM com os resultados do Oswestry, um instrumento bastante utilizado na população portuguesa, que mede o impacto da deficiência da dor lombar. Apesar de não existir um padrão-ouro para dor, segundo Scholtes et al., (2011) não é correto utilizar outro instrumento como padrão-ouro para estabelecer validade de critério.

A análise da confiabilidade demonstrou que 83% dos artigos (Kahraman et al., 2016; Namnik et al., 2015; Barros e Alexandre, 2003; Mesquita et al., 2010; Palmer et al., 1999) apresentaram qualidade metodológica pobre por não utilizarem um intervalo de tempo apropriado para reaplicar o QNSM. A maioria utilizou o tempo de uma semana, com exceção de Barros e Alexandre (2003) que aplicaram o questionário em

intervalo de um dia. Os autores justificaram a utilização desse intervalo para não alterar as respostas das variáveis estudadas. Entretanto, intervalo curto compromete a confiabilidade superestimando seu valor tendo como consequências resultados viesados (SCHOLTES et al., 2011).

Apenas Knudsen e colaboradores (2014) utilizaram o intervalo adequado entre as aplicações do QNSM, sendo este de quatro semanas. Segundo Scholtes e colaboradores (2011) o intervalo de tempo entre as medições deve ser suficientemente longo de forma que a primeira medição não afetará a segunda, sendo o intervalo ideal de duas a quatro semanas.

O que afetou a qualidade metodológica tanto nas pesquisas de validade quanto de confiabilidade, conforme o *checklist COSMIN* foi o item: Houve falhas na concepção ou metodologia do estudo? Segundo o *checklist COSMIN* um estudo recebe pontuação razoável para qualidade metodológica se houver uma pequena falha na sua metodologia ou concepção (Terwee et al., 2011). A falha considerada nos estudos em relação ao item citado foi o fato deles não deixarem evidência sobre a definição prévia do tamanho da amostra. Isso comprometeu a qualidade metodológica dos artigos analisados.

A falta de adequação estatística dos procedimentos amostrais no que se refere ao cálculo do tamanho da amostra foi um dos pontos críticos dos estudos analisados. É importante definir previamente o tamanho da amostra de uma pesquisa quantitativa, amostra muito grande causa gasto desnecessário de recursos, enquanto amostra muito pequena pode levar a resultados sem conclusão. A maioria dos artigos não apresentou os critérios utilizados no cálculo amostral. Apenas Kahraman e colaboradores (2016) descreveram tal critério e cálculo. Os autores fizeram cálculo da amostra para estudo de consistência interna e a utilizou também para avaliar a confiabilidade teste reteste.

Embora não tenha mencionado quais critérios foram utilizados para o cálculo do tamanho da amostra, Knudsen et al., (2014) e Namnik et al., (2015) relataram que utilizaram amostras por conveniência para selecionar os participantes da pesquisa. Esse tipo de amostra privilegia aqueles indivíduos com mais tempo disponível para preencher o questionário, enviesando os resultados.

Barros e Alexandre (2003) descreveram que a escolha dos participantes foi realizada de forma sistemática, porém eles não mencionam os critérios utilizados para o cálculo do tamanho da amostra. Os autores estudaram a confiabilidade numa amostra de quatro grupos distintos de profissionais utilizando o mesmo tamanho amostral nos subgrupos.

Nos estudos de Takekawa e colaboradores (2015) além de não mencionar os critérios para o cálculo, o tamanho final da amostra foi consequência da falta de tempo. A população era composta por 142 trabalhadores, sendo que 105 concordaram em participar da pesquisa. Porém, por falta de tempo, a amostra final do estudo foi composta por 50 trabalhadores. Pinheiro e colaboradores (2002) mencionam que o critério de escolha da amostra, a não aleatoriedade (auto-seleção), é responsável pelo viés, sendo uma das limitações do seu estudo. Enquanto Kahraman e colaboradores (2016) mencionam que ao calcular o tamanho da amostra, utilizaram o tipo de amostragem inadequada para realização do seu estudo. Os autores citam que ao invés de utilizar amostragem estratificada utilizaram amostragem aleatória simples para grupos heterogêneos de trabalhadores. Knudsen e colaboradores (2014) relatam que em seu estudo com 39 médicos cirurgiões ortopedistas residente, o tamanho da amostra é pequeno e pode não ser representativo para uma população dessa categoria de

trabalhadores. De acordo com o *COSMIN*, o tamanho da amostra no estudo de Knudsen e colaboradores (2014) é razoável.

Outro ponto crítico na avaliação dos estudos foi que os autores não levaram em consideração as mudanças ocorridas na ampliação do QNSM para melhorar sua especificidade. Todos os estudos sobre validade e/ou confiabilidade do questionário Nórdico avaliados aqui, utilizaram o questionário geral, versão de 1987, com exceção do estudo de Akesson et al., (1999) que utilizou o questionário específico para pescoço e ombro.

Verifica-se que os estudos não seguem um procedimento metodológico estruturado para avaliação da validade e da confiabilidade do QNSM. Namnik e colaboradores (2015), por exemplo, relatam que não avaliaram validade de critério por falta de um padrão-ouro para definir diagnóstico de DME. Embora não se tenha um padrão-ouro para definição de dor, o diagnóstico médico e o exame físico têm sido utilizados como padrão-ouro em vários estudos. Para avaliar um instrumento de triagem é imprescindível a avaliação da validade de critério, daí a necessidade de se analisar se a definição do padrão-ouro é adequada, apesar de não existir um padrão para se analisar dor.

É importante que os estudos de validade e confiabilidade sigam critérios metodológicos estabelecidos. O *checklist* COSMIN pode ser utilizado por pesquisadores para planejar estudos sobre propriedades de medidas de um instrumento (MOKKINK et al., 2010b). Mas em relação à amostragem o *checklist* COSMIN deixa uma lacuna, eles consideram uma amostra adequada se seu tamanho for superior a 100 indivíduos sem

dar importância à realização de um cálculo amostral, com critérios bem definidos tais como, nível de confiança, erro máximo estabelecido, variabilidade, etc.

Os resultados desta revisão corroboram com os de Anthonie e colaboradores (2014), que ao realizar uma revisão de literatura, com o objetivo de descrever as práticas utilizadas em estudos de validação psicométrica de instrumento de medidas utilizados na prática clínica. Concluíram que menos de 10% dos artigos analisados definiram os critérios utilizados para determinar o tamanho da amostra.

A maioria dos estudos examinados apresentou qualidade de medida para validade de critério sem classificação e qualidade metodológica razoável. Em relação a confiabilidade, a maioria dos estudos apresentou classificação indeterminada para qualidade de medida e qualidade metodológica pobre. Por fim, a qualidade geral das propriedades de medida validade de critério e confiabilidade teste-reteste do QNSM, para todas as regiões anatômicas, foi classificada como conflitante e desconhecida, respectivamente.

As limitações deste estudo se concentram quanto às questões do idioma e das propriedades de medida analisadas. Analisamos somente artigos que foram publicados em português e inglês. Desta forma seis artigos deixaram de fazer parte desta revisão por causa do idioma. No que se refere as propriedades de medida avaliadas neste estudo foram restritas a validade de critério e confiabilidade teste-reteste, alguns artigos avaliaram outros tipos de propriedades validade (transcultural, de conteúdo e de constructo) e a consistência interna.

Em função do constante uso do QNSM em todo mundo e dada sua importância enquanto instrumento de triagem para diagnóstico de DME, esta revisão enfatiza a

necessidade da realização de mais pesquisas com o objetivo de avaliar o QNSM em diversos grupos ocupacionais.

VI. 8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Akesson I, Johnsson B, Rylander L, Moritz U, Skerfving S. Musculoskeletal disorders among female dental personnel clinical examination and a 5-year follow-up study of symptoms. *Int Arch Occup Environ Health*, 72(6):395-403, 1999.
2. Anthoine E , Moret L ,Regnault A, Sbille V, Hardouin J B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, 12:176, 2014.
3. Barros ENC, Alexandre NMC. Cross-cultural adaptation of the Nordic Musculoskeletal Questionnaire. *Int Nurs Rev.*;50:101–108,2003.
4. Bjorkstén, MG, B Boquist, M Talback,C Edling The Validity of Reported Musculoskeletal Problems: a Study of Questionnaire Answers in Relation to Diagnosed Disorders and Perception of Pain. *Applied Ergonomics* 30, 325–330 1999.
5. Baron S, HalesT, HurreL J. Evaluation of symptom surveys for occupational musculoskeletal disorders.*American Journal of Industrial Medicine*, 29: 609-17, 1996
6. Risk factors for work-related musculoskeletal disorders: a systematic review of recente Costa BR, Vieira ER longitudinal studies. *Amer J Industr Med* 53:285-323, 2010.
7. Descatha A, Roquelaure Y, Chastang JF, Evanoff B, Melchior M, Mariot C, HA C, Imbernon E, Goldberg M, Leclerc A. Validity of Nordic-style questionnaires in the surveillance of upper-limb work-related musculoskeletal disorders. *Scand J Work Environ Health*, v. 1, n. 33, p. 58-65, 2007.
8. Elbers RG, Rietberg M B, van Wegen E E, Verhoef J, Kramer, SF, Terwee CB, Kwakkel G. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of Life Research*, v. 21, n. 6, p.925-944, 2012.
9. Fernandes RCP, Assunção AA, Silvany Neto AM, Carvalho FM. Musculoskeletal disorders among workers in plastic manufacturing plants. *Revista Brasileira de Epidemiologia*, São Paulo, 13 (1): 11-20, 2010.

10. Kahraman T, Genç A, Goz E. The Nordic Musculoskeletal Questionnaire: cross-cultural adaptation into Turkish assessing its psychometric properties. *Disability and Rehabilitation*, 2015.
11. Knudsen ML, PMLudewig, Jonathan PB. Musculoskeletal pain in resident orthopaedic surgeons: results of a novel survey. *The Iowa orthopaedic journal*, v.34,190-196,2014.
12. Kuorinka I, Forcier L. Work related musculoskeletal disorders (WMSDs): a reference book for prevention. London. Taylor e Francis, p.213-245, 1995.
13. Kuorinka I, Jonsson B, Kilbom a, Vinterberg H, Biering-sorensen f, Andersson G. et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon*, v.18,233-237, 1987.
14. Mesquita CC, Ribeiro JC, Moreira P. Portuguese version of the standardized Nordic Musculoskeletal Questionnaire: cross cultural and reliability. *J Public Health (Oxf)*.;18:461-466,2010.
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, v. 19, n. 4, p. 539-549, 2010a.
16. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, v. 63, N. 7, P. 737 - 745, 2010b.
17. Moher D, Liberti A, Telzla J, Altam DG. Principais itens para relatar Revisões Sistemáticas e Meta-análises: A recomendação PRISMA. Tradução: Taís Feire Galvão e Thaís de Souza Andrade Pansani; retro-traduzido por: David Harrad. *Epidemiol. Serv. Saúde*, Brasília 24(2): abr-jun 2015.
18. Monteiro, G. T. R. Pesquisa em Saúde Pública: como desenvolver e validar instrumentos de coleta de dados. Curitiba: Appis, 2014.
19. Namnik N, H Negahban, R Salehi, R Shafizadeh, MS Tabib. Validity and reliability of Persian version of the Specific Nordic questionnaire in Iranian industrial workers. *Work*, v.54,35-41,2016.
20. Ohlsson K, Attewell RG, Alm A, Johnsson B, Skerfving S. An assessment of neck and upper extremity disorders by questionnaire and clinical examination. *Ergonomics*, v.37, 891-897,1994.

21. Punnett L and Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyography Kinesiol.*14:13-23, 2004.
22. Palmer K, Smith G, Kellingray S, Cooper C. Repeatability and validity of an upper limb and neck discomfort questionnaire: the utility of the standardized Nordic questionnaire. *Occup. Med.*,v. 49, n. 3, p. 171-5. 1999.
23. Pinheiro FA, Tróccolia BT, Carvalho CV. Validação do Questionário Nórdico de Sintomas Osteomusculares como medida de morbidade. *Revista de Saúde Pública*, São Paulo, v. 36, n. 3, p. 307-12. 2002.
24. Perreault N, Brisson C, Dionne CE, Montreuil S, Punnet L. Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck-shoulder region and a physical examination. *BMC Musculoskeletal Disorders*, v. 9, n. 34. 2008.
25. Stal M, Moritz U, Johnsson B, Pinzke S . The natural course of musculoskeletal symptoms and clinical findings in upper extremities of female milkers. *Int J Occup Environ Health*, v.3, 190-197,1997.
26. Silverstein BA, Stetson DS, Keyserling WM, Fine LJ. Work-related musculoskeletal disorders: comparison of data sources for surveillance. *Am J Ind Med*, v. 31, p. 600–8. 1997.
27. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury*,v.42,236-240,2011.
28. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, v. 19, n. 4, p. 539-549, 2012.
29. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, v. 60, n. 1, p. 34-42, 2007.
30. Takekawa KS, Gonçalves JS, Moriguchi CS, Coury HJCG, Sato TO. Can a self-administered questionnaire identify workers with chronic or recurring low backpain? *Industrial Health*,v. 53, 340 – 345. 2015.

VI. 9. TABELAS

Tabela 1: Critérios para avaliação da qualidade das propriedades de medida.

Propriedades	Critério de Qualidade^a
Validade de critério	<p>(+) Argumentos convincentes de que o padrão-ouro é “ouro” e correlação com o padrão-ouro $\geq 0,70$.</p> <p>(?) Não há argumentos convincentes de que o padrão-ouro é “ouro” ou a metodologia duvidosa.</p> <p>(-) Correlação com o padrão-ouro $< 0,70$, apesar da concepção e método adequado.</p> <p>(0) Não há informações sobre a validade de critérios</p>
Confiabilidade teste-reteste	<p>(+) CCI (Coeficiente de correlação intraclasse) ou kappa $\geq 0,70$.</p> <p>(?) Planejamento ou método duvidoso (ex. intervalo de tempo entre as medidas não mencionadas).</p> <p>(-) CCI ou kappa $< 0,70$, apesar da concepção e método adequado.</p> <p>(0) Não há informações sobre a confiabilidade.</p>

Fonte: Terwee et al., (2007) . (Tradução e adaptação da autora.)

(+) = adequada; (-) = não adequada; (?) = indeterminada; (0) = ou sem classificação

^aPlanejamento ou Metodologia do estudo duvidosa = falta de uma clara descrição do planejamento ou método do estudo, tamanho de amostra menor que 50 indivíduos (deve ser de no mínimo 50 em cada subgrupo de análise) ou qualquer falha metodológica importante no planejamento ou execução do estudo.

Tabela 2: Critério de avaliação da qualidade metodológica dos estudos de acordo com a escala de classificação de quatro pontos por itens do checklist COSMIN.

Itens	Escala de classificação
Validade de critério	
1.A percentagem de itens faltantes foi dada? sim <input type="checkbox"/> não <input type="checkbox"/>	Porcentagem de itens faltantes foi descrita. Excelente . Porcentagem de itens faltantes não foi descrita. Boa .
2.Houve uma descrição de como os itens faltantes foram manipulados? sim <input type="checkbox"/> não <input type="checkbox"/>	Descreveu como os itens faltantes foram tratados – Excelente . Não descrito, mas pode ser deduzido como os itens faltantes foram tratados – Boa . Não está claro como os itens faltantes foram tratados – Fraca .
3.O tamanho da amostra incluído na análise foi adequado? sim <input type="checkbox"/> não <input type="checkbox"/>	Amostra adequada tamanho (≥ 100). Excelente Tamanho da amostra bom (50-99). Boa Amostra moderada tamanho (30-49). Razoável Amostra de tamanho pequeno (< 30). Pobre
4.O critério utilizado ou empregado pode ser considerado como um padrão-ouro razoável? sim <input type="checkbox"/> não <input type="checkbox"/>	Critério utilizado pode ser considerado um padrão-ouro adequado (há evidência). Excelente Nenhuma evidência, mas é possível assumir que o critério utilizado pode ser considerado um padrão-ouro adequado. Boa Não está claro se o critério utilizado pode ser considerado padrão-ouro adequado. Razoável Critério utilizado não pode ser considerado um padrão-ouro adequado. Pobre
5.Houve alguma falha importante na concepção ou no método do estudo? sim <input type="checkbox"/> não <input type="checkbox"/>	Não houve nenhuma falhas na metodologia ou na execução do estudo. Excelente Houve pequenas falhas na metodologia ou na execução do estudo. Razoável Houve importantes falhas na metodologia ou na execução do estudo. Pobre
6.Para pontuações contínuas: Foram correlações, ou a área sob o Receptor Operando Curva (curva ROC) calculada? sim <input type="checkbox"/> não <input type="checkbox"/>	Correlações ou curva ROC calculada. Excelente Correlações ou curva ROC não calculados. Pobre
7.Para pontuações dicotômicas: Foram determinadas a sensibilidade e a especificidade? sim <input type="checkbox"/> não <input type="checkbox"/>	Sensibilidade e Especificidade calculadas. Excelente Sensibilidade e Especificidade não calculadas. Pobre

Confiabilidade teste-reteste	
1.O percentual de itens faltantes foi descrito? sim <input type="checkbox"/> não <input type="checkbox"/>	Percentual de itens faltantes descrito. Excelente Percentual de itens faltantes não descrito. Boa
2.Houve uma descrição de como os itens faltantes foram tratados? sim <input type="checkbox"/> não <input type="checkbox"/>	Descreveu como itens faltantes foram tratados. Excelente Não descrito, mas pode ser deduzido como os itens faltantes foram tratados. Boa Não está claro como itens faltantes foram tratados. Razoável
3.O tamanho da amostra foi adequado ao estudo? sim <input type="checkbox"/> não <input type="checkbox"/>	Tamanho de amostra adequado (≥ 100). Excelente Tamanho da amostra bom (50-99). Boa Tamanho de amostra moderado (30-49). Razoável Tamanho de amostra pequeno (< 30). Pobre
4.Pelo menos duas medições foram realizadas? sim <input type="checkbox"/> não <input type="checkbox"/>	Ao menos duas medições foram realizadas. Excelente Apenas uma medição foi realizada. Pobre
5.As medições foram realizadas de modo independente? sim <input type="checkbox"/> não <input type="checkbox"/>	Medições independentes. Excelente Assumindo que as medições foram independentes. Boa Duvidoso se as medições foram independentes. Razoável Medições não independentes. Pobre
6.O intervalo de tempo entre medições foi indicado? sim <input type="checkbox"/> não <input type="checkbox"/>	Declarou o intervalo de tempo. Excelente Intervalo de tempo não declarado. Razoável
7.Os pacientes permaneceram estáveis no período entre medições em relação ao construto a ser medido? sim <input type="checkbox"/> não <input type="checkbox"/>	Os pacientes ficaram estáveis (evidência fornecida). Excelente Assumindo que pacientes ficaram estáveis. Boa Não está claro se os pacientes ficaram estáveis. Razoável Os pacientes não ficaram estáveis. Pobre
8.O intervalo de tempo entre medições foi adequado? sim <input type="checkbox"/> não <input type="checkbox"/>	Intervalo de tempo apropriado Excelente Duvidoso se intervalo de tempo foi apropriado. Razoável Intervalo de tempo não apropriado. Pobre
9.As condições de teste foram semelhantes para ambas as medições? Tipo de administração, ambiente, instruções sim <input type="checkbox"/> não <input type="checkbox"/>	Condições do teste eram semelhantes (evidência forneceu). Excelente Assumindo que as condições de teste eram semelhantes. Boa Não está claro se o teste as condições foram semelhantes. Razoável Condições do teste não eram semelhantes. Pobre
10.Houve alguma falha importante no desenho ou métodos do estudo? sim <input type="checkbox"/> não <input type="checkbox"/>	Não houve nenhuma falha na metodologia ou na execução do estudo. Excelente Houve pequenas falhas na metodologia ou na execução do estudo. Razoável Houve importantes falhas na metodologia ou na execução do estudo. Pobre

11. Para pontuações contínuas: Foi um coeficiente de correlação intraclasse (ICC) calculado?	ICC calculado e modelo ou Fórmula do ICC descrito. Excelente ICC calculado mas modelo ou fórmula do ICC não descrito ou não otimizado. Coeficiente de correlação de Pearson ou Spearman calculado com a evidência, desde que não tenha ocorrido mudança sistemática. Boa Coeficiente de correlação de Pearson ou Spearman calculado sem evidência desde que nenhuma mudança sistemática tenha ocorrido ou com evidência de que ocorreu mudança sistemática. Razoável Nenhum ICC ou Pearson ou coeficiente de correlação de Spearman calculado. Pobre
12. Para pontuações dicotômicas / nominais / ordinal: Foi calculado kappa? sim <input type="checkbox"/> não <input type="checkbox"/> NA <input type="checkbox"/>	Kappa calculado. Excelente Apenas porcentagem de concordância calculado. Pobre
13. Para pontuações ordinais: Foi calculado um kappa ponderado? sim <input type="checkbox"/> não <input type="checkbox"/> NA <input type="checkbox"/>	Kappa ponderado calculado. Excelente Kappa ponderado não calculado. Razoável Apenas porcentagem de concordância calculado.
14. Para pontuações ordinais: Foi descrito o esquema de ponderação? Por exemplo, linear, quadrático sim <input type="checkbox"/> não <input type="checkbox"/> NA <input type="checkbox"/>	Esquema de ponderação descrito. Excelente Esquema de ponderação não descrito. Boa

Adaptado e traduzido pela autora, retirado do *checklist* COSMIN e COSMIN escala de classificação de quatro pontos disponível em: http://www.cosmin.nl/cosmin_checklist.html. Acesso em 11 de junho de 2017

Tabela 3: Critérios para determinar a qualidade geral das propriedades de medida de um instrumento.

Classificação	Avaliação obtida pela propriedade de medida	Critérios para determinação da qualidade geral das propriedades de medidas de um instrumento
Forte	"Adequado" ou "Não adequado" (+ ou -)	Achados consistentes em múltiplos estudos de boa qualidade metodológica ou em estudo de qualidade metodológica "excelente"
Moderado	"Adequado" ou "Não adequado" (+ ou -)	Resultados consistentes em múltiplos estudos de qualidade metodológica "razoável" ou em estudo de "boa" qualidade metodológica
Limitado	"Adequado" ou "Não adequado" (+ ou -)	Um estudo de qualidade metodológica "razoável"
Conflitante	"Conflitante" (\pm)	Achados conflitantes
Desconhecido	'Desconhecido' (?)	Apenas estudos de qualidade metodológica "pobre"

Fonte: Elbers et al., 2012. (Tradução e adaptação da autora).

Tabela 4: Características dos estudos de validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesquelético.

Autores, ano; País	Título	Propriedades de medida avaliadas	Grupo estudado	Definição de casos		Versão do QNSM utilizada
				Sintomas no QNSM	Padrão-ouro	
Kahraman et al., 2016; Turquia	The Nordic Musculoskeletal Questionnaire: cross-cultural adaptation into Turkish assessing its psychometric properties	Confiabilidade teste-reteste	198 profissionais de saúde de um hospital universitário	Dor, dor e desconforto nos últimos 12 meses e 7 dias; e Impedimento de realizar tarefas diárias nos últimos 12 meses por causa do problema.		Questionário geral versão de 1987
Namnik et al., 2015; Irã	Validity and reliability of Persian version of the Specific Nordic questionnaire in Iranian industrial workers	Confiabilidade teste-reteste	122 trabalhadores da indústria (60 participaram do estudo de confiabilidade).	Dor, dor e desconforto nos últimos 7 dias		Questionário geral versão de 1987
Takekawa et al., 2015; Brasil	Can a self-administered questionnaire identify workers with chronic or recurring low back pain?	Validade	50 trabalhadores de departamento administrativo de uma universidade pública	Dor, formigamento ou dormência nos últimos 12 meses e últimos 7 dias; Limitações funcionais para realizar atividades laborais, domésticas e de lazer devido à dor; Procura por cuidados médicos por causa da dor.	Avaliação física que inclui teste de mobilidade, força, palpação e testes neurológicos.	Questionário geral versão de 1987
Knudsen et al., 2014; EUA	Musculoskeletal pain in resident orthopaedic surgeons: results of a novel survey	Confiabilidade	39 residentes de cirurgia ortopédica (15 participaram do reteste).	Dor, dor e desconforto nos últimos 12 meses; Dor, rigidez, fraqueza ou parestesia nos últimos 7 dias; Impedimento para realização das atividades rotineira nos últimos 12 meses devido aos sintomas.		Questionário geral versão de 1987

Autores, ano; País	Titulo	Propriedades de medida avaliadas	Grupo estudado	Definição de casos		Versão do QNSM utilizada
				Sintomas no QNSM	Padrão-ouro	
Mesquita et al.,2010; Portugal	Portuguese version of the standardized Nordic musculoskeletal questionnaire: Cross cultural and reliability	Validade e Confiabilidade e teste-reteste	60 trabalhadores de um armazém de alimentos	Dor, dor e desconforto, e dormência nos últimos 12 meses e 7 dias; e Impedimento de realizar tarefas diárias por causa do problema.	Sintomas avaliados através do Oswestry, questionário que mede o impacto de deficiência da dor lombar.*	Questionário geral versão de 1987
Perreault et al.,2008; EUA	Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck-shoulder region and a physical examination	Validade	187 trabalhadores de escritório	Sintoma durante três dias nos últimos 7 dias, ou intensidade de dor ≥ 50 milímetros numa escala visual analógica de 100 milímetros.	Diminuição $\geq 30\%$ da amplitude ativa de movimento. Diminuição da força muscular normal pontuação ≤ 4 na escala de Lovett Escala de força muscular). Dor com intensidade ≥ 3 na escala de 11 pontos no local relevante durante qualquer manobra.	Questionário geral versão de 1987
Descatha et al.,2007; França	Validity of Nordic-style questionnaires in the surveillance of upper-limb work-related musculoskeletal disorders	Validade	2.355 trabalhadores expostos a trabalho repetitivo (1757 trabalhadores em 1993-1994 e 598 trabalhadores em 1996-1997); 2685 trabalhadores em 2002-2003 que faziam parte de um sistema de vigilância	Presença de pelo menos um sintoma relacionada à dor e desconforto nos últimos 6 meses considerando todos os distúrbios no exame clínico e análise restrita a seis distúrbios; Presença de pelo menos um sintoma relacionada à dor e desconforto nos últimos 12 meses.	Exame clínico padronizado baseado em um uma lista de critérios para diagnóstico; Exame clínico padronizado baseado em um protocolo internacional para avaliação de DME em membros superiores.	Questionário geral versão de 1987

Autores, ano; País	Titulo	Propriedades de medida avaliadas	Grupo estudado	Definição de casos		Versão do QNSM utilizada
				Sintomas no QNSM	Padrão-ouro	
Barros e Alexandre,2003; Brasil	Cross-cultural adaptation of the Nordic musculoskeletal questionnaire	Confiabilidade teste-reteste	40 indivíduos distribuídos em 4 grupos distintos (alunos de graduação em enfermagem, professores do departamento de enfermagem, funcionários administrativos da universidade e pacientes atendidos pelo serviço de fisioterapia do serviço médico universitário)	Dor, formigamento, dormência nos últimos 12 meses e nos últimos 7 dias; Limitações para realização de tarefas diárias por causa da dor e Procura por atenção médica devido ao problema.	-----	Questionário geral versão de 1987
Pinheiro et al.,2002; Brasil	Validity of the Nordic Musculoskeletal Questionnaire as morbidity measurement tool	Validade	90 bancários	Sintomas nos últimos: 12 meses ou 7 dias; 12 meses e 7 dias; 12 meses ou 7 dias e absenteísmo; 12 meses e 7 dias e absenteísmo.	Diagnóstico realizado por um médico do trabalho através da análise da história clínica e diagnóstico do prontuário do trabalhador na empresa.	Questionário geral versão de 1987
Palmer et al.,1999; Brasil	Repeatability and validity of an upper limb and neck discomfort questionnaire: The utility of the standardized Nordic questionnaire	Validade e Confiabilidade teste-reteste	105 pacientes ambulatoriais	Dor com duração de um dia ou mais na semana passada, e no ano anterior; que tenha, impedido a realização de atividades diárias normais.	Diagnóstico realizado por um reumatologista de espondilose cervical, capsulite adesiva e epicondilite lateral	Questionário geral versão de 1987

Autores, ano; País	Titulo	Propriedades de medida avaliadas	Grupo estudado	Definição de casos		Versão do QNSM utilizada
				Sintomas no QNSM	Padrão-ouro	
Akesson et al.,1999; Suécia	Musculoskeletal disorders among female dental personnel - Clinical examination and a 5-year follow-up study of symptoms	Validade	90 profissionais de odontologia	Dor e desconforto nos últimos 12 meses e 7 dias anterior a entrevista	Diagnóstico de exame físico realizado de acordo com um protocolo padronizado	Questionário geral versão de 1987
Bjorksten et al.,1998; Suécia	The validity of reported musculoskeletal problems. A study of questionnaire answers in relation to diagnosed disorders and perception of pain	Validade	171 trabalhadoras (empresa de processamento de alimentos e serviço industrial)	Dor e desconforto nos últimos 3 meses e 7 dias	Diagnóstico realizado através de uma avaliação clínica, com critérios predefinidos utilizado no exame clínico.	Questionário geral versão de 1987
Bjorksten et al.,1998; Suécia	The validity of reported musculoskeletal problems. A study of questionnaire answers in relation to diagnosed disorders and perception of pain	Validade	171 trabalhadoras (empresa de processamento de alimentos e serviço industrial)	Dor e desconforto nos últimos 3 meses e 7 dias	Diagnóstico realizado através de uma avaliação clínica, com critérios predefinidos utilizado no exame clínico.	Questionário geral versão de 1987
Silverstein et al., 1997; Canadá	Work-Related Musculoskeletal Disorders: Comparison of Data Sources for Surveillance	Validade	626 trabalhadores de indústria	Sintoma pelo menos 20 vezes durante o último ano. Com duração de no mínimo 1 semana.	Não exigiu limitação de movimento resistido; Dor no local da queixa durante a manobra, agravada por testes de movimento resistido.	Questionário geral versão de 1987

Autores, ano; País	Titulo	Propriedades de medida avaliadas	Grupo estudado	Definição de casos		Versão do QNSM utilizada
				Sintomas no QNSM	Padrão-ouro	
Stal et al., 1997; Suécia	The natural course of musculoskeletal symptoms and clinical findings in upper extremities of female milkers	Validade	161 trabalhadores de ordenha do sexo feminino	Dor, dor e desconforto nos últimos 12 meses e 7 dias	Diagnóstico clínico definido através de resultados de um exame físico com critérios padronizados	Questionário geral versão de 1987
Ohlsson et al., 1994; Suécia	An assessment of neck and upper extremity disorders by questionnaire and clinical examination	Validade	165 mulheres trabalhadoras de uma fábrica em ocupações variadas	Dor, dor ou desconforto nos últimos 7 dias.	Diagnóstico clínico definido através de resultados de um exame físico com critérios padronizados	Questionário geral versão de 1987

Tabela 5: Metodologia utilizada e região anatômica avaliada nos estudos selecionados.

Autor; Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Kahraman et al., 2016; 198 profissionais de saúde de um hospital universitário	Confiabilidade teste-reteste avaliada através do coeficiente kappa com intervalo de 1 semana entre as aplicações do QNSM.	k = 0,74 ^a k = 0,75 ^b	k = 0,68 ^a k = 0,61 ^b	k = 0,89 ^a k = 0,88 ^b	k = 0,57 ^a k = 0,90 ^b	k = 0,73 ^a K = 0,63 ^b	k = 0,78 ^a k = 0,74 ^b	k = 0,79 ^a k = 0,74 ^b	k = 0,71 ^a k = 0,72 ^b	k = 0,73 ^a k = 0,71 ^b		
Namnik et al., 2015; 122 trabalhadores da indústria (60 participaram do estudo de confiabilidade).	Confiabilidade teste-reteste avaliada através do coeficiente kappa com intervalo de 1 semana entre as aplicações do QNSM.				k = 0,83 ^a				k = 0,71 ^a			
Takekawa et al., 2015; 50 trabalhadores do departamento administrativo de uma universidade pública	Validade critério avaliada através da sensibilidade, especificidade com combinações de 3 métodos de avaliação incluindo o QNSM e exame físico.				S = 100% ^a E = 88% ^a							

Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Knudsen et al., 2014; 39 residentes de cirurgia ortopédica.	Confiabilidade teste-reteste avaliada através do coeficiente kappa com intervalo de 4 semanas entre as aplicações do QNSM.	k = 0,19 ^a k = 0,84 ^b	k = 0,33 ^a k = 0,60 ^b		k = - ^a k = 0,84 ^b	k = -0,07 ^a k = 0,73 ^b	k = 0,40 ^a k = 0,81 ^b	k = 1,0 ^a k = 1,0 ^b	k = -0,80 ^a k = 0,72 ^b	k = - ^a k = 0,63 ^b		
Mesquita et al., 2010; 60 trabalhadores de um armazém de alimentos.	Confiabilidade teste-reteste avaliada através do coeficiente kappa com intervalo de 1 semana entre as aplicações do QNSM; Validade concorrente avaliada através do coeficiente de correlação tau-de Kendal.	k = 0,88 ^a k = 0,67 ^b	k = 1,0 ^a k = 1,0 ^b		k = 0,93 ^a k = 0,86 ^b C = 0,29 ^a C = 0,47 ^a	= 0,93 ^a k = 0,78 ^b C = 0,35 ^a	k = 1,0 ^b k = 1,0 ^b		k = 1,0 ^b	k = 1,0 ^b		
Perreault et al., 2008; 187 trabalhadores de escritório	Validade critério avaliada através da sensibilidade, especificidade e concordância entre o QNSM e o exame físico utilizando o kappa											S = 75% ^a S = 56% ^a E = 80% ^a E = 78% ^a

Autor; Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Descatha et al.,2007; 2.355 trabalhadores expostos a trabalho repetitivo (1757 trabalhadores em 1993-1994 e 598 trabalhadores em 1996-1997) 2685 trabalhadores em 2002-2003 que faziam parte de um sistema de vigilância	Validade critério avaliada através do kappa, sensibilidade, especificidade e curva ROC.										1993-1994 S = 94% ^a E = 81,4% ^a 1996-1997 S = 82,3% ^a E = 77,7% ^a 2002-2003 S = 100% ^a E = 51,1% ^a Validade concorrente 1993-1994 k= 0,77 ^a 1996-1997 k= 0,57 ^a 2002-2003 k= 0,22 ^a	

Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Barros e Alexandre, 2003; 40 indivíduos distribuídos em 4 grupos distintos	Confiabilidade teste-reteste avaliada através do coeficiente kappa com intervalo de 1 dia entre as aplicações do QNSM.	k = 0,78 ^a k = 0,67 ^b	k = 0,76 ^a k = 0,85 ^b	k = 0,48 ^a k = 1,0 ^b	k = 0,82 ^a k = 0,75 ^b	k = 0,63 ^a k = 1,0 ^b	k = 0,80 ^a k = 0,78 ^b	k = 1,0 ^a k = 0,83 ^b	k = 0,78 ^a k = 0,84 ^b	k = 1,0 ^a k = 0,90 ^b		
Pinheiro et al., 2002; 90 bancários	Validade concorrente avaliada através da correlação de Pearson entre sintomas e afastamentos (história clínica) por sexo e região anatômica.		CP = 0,60 ^a CP = 0,32 ^b	CP = 0,79 ^a CP = 0,67 ^b	CP = 0,50 ^a CP = 0,62 ^b		CP = 0,71 ^a CP = 0,61 ^b	CP = 0,36 ^a CP = 0,49 ^b	CP = 0,44 ^a CP = 0,46 ^b	CP = 0,39 ^a CP = - ^b		

Autor; Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Palmer et al.,1999; 105 pacientes ambulatoriais	Validade de critério e confiabilidade, a validade foi avaliada através da sensibilidade e especificidade. Enquanto a confiabilidade pelo índice kappa com intervalo de 1 semana entre as aplicações do QNSM.	k = 0,64 ^a k = 0,75 ^b E= 90% ^b	k = 0,71 ^a k = 0,74 ^b E= 100% ^b	k = 0,69 ^a k = 0,82 ^b E= 90% ^b			k = 0,68 ^a k = 0,73 ^b E= 100% ^b E= 78% ^b					
Akesson et al.,1999; 90 profissionais de odontologia	Validade critério avaliada através da sensibilidade, especificidade.	S= 75% ^a S= 100% ^b E=72% ^a E= 47% ^b	S= 74% ^a S= 95% ^b E = 60% ^a E= 48% ^b	S = 71% ^a S= 71% ^b E = 91% ^a E=84% ^b			S= 62% ^a S= 62% ^b E = 78% ^a E=64% ^b	S= 78% ^a S= 89% ^t E=93% ^a E= 87% ^t				
Bjorksten et al.,1998; 171trabalhadoras (empresa de processamento de alimentos e serviço industrial)	Validade de critério avaliada através da sensibilidade e especificidade.	S = 92% ^a E = 62% ^a	S = 100% ^a E= 41% ^a									

Autor; Grupo estudado	Metodologia utilizada	Região anatômica avaliada										
		Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Pulsos/mãos	Quadril / Coxa	Joelhos	Tornozelos / pé	M.S*	R.C**
Silverstein et al., 1997; 626 trabalhadores de indústria	Validade preditiva e concorrente	S= 75% ^a S= 100% ^b E = 72% ^a E= 47% ^b	S= 74% ^a S= 95% ^b E = 60% ^a E= 48% ^b	S = 71% ^a S= 71% ^b E = 91% ^a E= 84% ^b			S = 62% ^a S= 62% ^b E= 78% ^a E= 64% ^b					
Stal et al., 1997; 161 trabalhadores de ordenha do sexo feminino	Validade de critério avaliada através da sensibilidade e especificidade.	S = 58% ^a E = 58% ^a	S = 52% ^a E = 86% ^a	S = 60% ^a E = 98% ^a			S = 53% ^a E = 96% ^a					
Ohlsson et al., 1994; 165 mulheres trabalhadoras de uma fábrica em ocupações variadas	Validade de critério avaliada através da sensibilidade e especificidade.	S = 66% ^a E = 84% ^a	S = 92% ^a E = 71% ^a	S = 79% ^a E = 88% ^a			S = 67% ^a E = 76% ^a					

*= Membros superiores (ombros, braços, cotovelos, antebraços, punho e mão); ** = Região Cervical (pescoço / ombros) ; k = kappa coeficiente de concordância; S= sensibilidade; E= especificidade; a = sintomas (problemas) nos últimos 7 dias; b= sintomas (problemas) nos últimos 12 meses; c= Correlação tau-de Kendall; CP= coeficiente de correlação de Pearson.

Tabela 6: Avaliação da qualidade de medida validade de critério do Questionário Nórdico de Sintomas Musculoesqueléticos por regiões anatômicas.

Autores	Regiões anatômicas avaliadas- validade de critério										
	Pescoço	Ombros	Cotovelos	Região lombar	Parte alta das costas	Punhos/ mãos	Quadril /Coxa	Joelhos	Tornozelos/ pé	M.S*	R.C**
Takekawa et al., 2015				0							
Mesquita et al., 2010				?	?						
Perreault et al., 2008.											0
Descatha et al., 2007										0	
Pinheiro et al., 2002		?	?	?		?	?	?	?		
Palmer et al., 1999	0	0	0			0					
Akesson et al., 1999	0	0	0			0	0				
Bjorksten et al., 1998	0	0									
Silverstein et al., 1997	0	0	0			0					
Stal et al., 1997	0	0	0			0					
Ohlsson et al., 1994	0	0	0			0					

(+) = adequada; (-) = não adequada; (?) = indeterminada; (0) = sem classificação; *= Membros superiores (ombros, braços, cotovelos, antebraços, punho e mão); ** = Região Cervical (pescoço / ombros).

Tabela 7: Avaliação da qualidade de medida confiabilidade teste-reteste do Questionário Nórdico de Sintomas Musculoesqueléticos por regiões anatômicas.

Autores	Regiões anatômicas avaliadas-confiabilidade teste-reteste								
	Pescoço	Ombro	Cotovelos	Região lombar	Parte alta das costas	Punhos /mãos	Quadril /Coxa	Joelhos	Tornozelos /pé
Kahraman et al.,2016	?	?	?	?	?	?	?	?	?
Namnik et al., 2015				?				?	
Knudsen et al., 2014	+	-	+	+	+	+	+	+	-
Mesquita et al., 2010	?	?		?	?	?		?	?
Barros e Alexandre, 2003	?	?	?	?	?	?	?	?	?
Palmer et al., 1999	?	?	?			?			

(+) = adequada; (-) = não adequada; (?) = indeterminada; (0) = sem classificação.

Tabela 8: Avaliação da qualidade metodológica dos estudos de validade e confiabilidade do Questionário Nórdico de Sintomas Musculoesqueléticos (QNSM) através do checklist COSMIN.

Autores	Propriedade de medida avaliada	
	Validade de critério	Confiabilidade teste-reteste
Kahraman et al., 2016		Pobre
Namnik et al., 2015		Pobre
Takekawa et al., 2015	Razoável	
Knudsen et al., 2014		Razoável
Mesquita et al., 2010	Pobre	Pobre
Perreault et al., 2008.	Razoável	
Descatha et al., 2007	Razoável	
Barros e Alexandre, 2003		Pobre
Pinheiro et al., 2002	Razoável	
Palmer et al., 1999	Excelente	Pobre
Akesson et al., 1999	Razoável	
Bjorksten et al., 1998	Razoável	
Silverstein et al., 1997	Razoável	
Stal et al., 1997	Razoável	
Ohlsson et al., 1994	Razoável	

VII. LISTA GERAL DE REFERÊNCIAS BIBLIOGRÁFICAS

1. Akesson I, Johnsson B, Rylander L, Moritz U, Skerfving S: Musculoskeletal disorders among female dental personnel—clinical examination and a 5-year follow-up study of symptoms. *Int Arch Occup Environ Health*, 72(6):395-403 1999.
2. Anthoine E , Moret L ,Regnault A, Sбилle V, Hardouin J B. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, 12:176, 2014.
3. Barros ENC, Alexandre NMC. Cross-cultural adaptation of the Nordic Musculoskeletal Questionnaire. *Int Nurs Rev.*;50:101–108,2003.
4. Bjorkstén, MG, B Boquist, M Talback,C Edling The Validity of Reported Musculoskeletal Problems: a Study of Questionnaire Answers in Relation to Diagnosed Disorders and Perception of Pain. *Applied Ergonomics* 30, 325–330 1999.
5. Baron S, HalesT, HurreL J. Evaluation of symptom surveys for occupational musculoskeletal disorders.*American Journal of Industrial Medicine*, 29: 609-17, 1996
6. Costa BR, Vieira ER. Risk factors for work-related musculoskeletal disorders: a systematic review of recente longitudinal studies. *Amer J Industr Med* 53:285-323, 2010.
7. Descatha A, Roquelaure Y, Chastang JF, Evanoff B, Melchior M, Mariot C, HA C, Imbernon E, Goldberg M, Leclerc A. Validity of Nordic-style questionnaires in the surveillance of upper-limb work-related musculoskeletal disorders. *Scand J Work Environ Health*, v. 1, n. 33, p. 58-65, 2007.
8. Elbers RG, Rietberg M B, van Wegen E E, Verhoef J, Kramer, SF, Terwee CB, Kwakkel G. Self-report fatigue questionnaires in multiple sclerosis, Parkinson’s disease and stroke: a systematic review of measurement properties. *Quality of Life Research*, v. 21, n. 6, p.925-944,2012.
9. EUA. Department of Labor. Nonfatal Occupational Injuries and Illnesses Requiring Days Away From Work, 2015. Disponível em: <https://www.bls.gov/news.release/osh2.nr0.htm> . Acesso em: 12 de fevereiro de 2017.
10. Fernandes RCP, Pataro SMS , Carvalho RB, Burdor A. The concurrence of musculoskeletal pain and associated work-related factors: a cross sectional study. *BMC Public Health*,v.16; 2016

11. Fernandes RCP, Assunção AA, Carvalho FM. Musculoskeletal disorders among workers in plastic industry. *Caderno de saúde pública*, Rio de Janeiro, 27(1), 13 (1): 78-86, 2011.
12. Fullerton JT. Evaluation of research studies. Part IV: Validity and reliability-- concepts and application. *J Nurse Midwifery*, V.38,121-125,1993.
13. INSS, Instituto Nacional de Seguro Social. Auxílios-doença acidentários e previdenciários concedidos segundo os códigos da Classificação Internacional de Doenças–CID-10. Disponível em: <http://www.previdencia.gov.br/dados-abertos/estatsticas/tabelas-cid-10/>. Acesso em: 18 de fevereiro de 2017.
14. Kahraman T, Genç A, Goz E. The Nordic Musculoskeletal Questionnaire: cross-cultural adaptation into Turkish assessing its psychometric properties. *Disability and Rehabilitation*, 2015.
15. Knudsen ML, PMLudewig, Jonathan PB. Musculoskeletal pain in resident orthopaedic surgeons: results of a novel survey. *The Iowa orthopaedic journal*, v.34,190-196,2014.
16. Kuorinka I, Forcier L. Work related musculoskeletal disorders (WMSDs): a reference book for prevention. London. Taylor e Francis, p.213-245, 1995.
17. Kuorinka I, Jonsson B, Kilbom a, Vinterberg H, Biering-Sorensen f, Andersson G. et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon*, v.18,233-237, 1987.
18. Mesquita CC, Ribeiro JC, Moreira P. Portuguese version of the standardized Nordic Musculoskeletal Questionnaire: cross cultural and reliability. *J Public Health (Oxf)*.;18:461–466,2010.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, v. 19, n. 4, p. 539-549, 2010.
20. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, v. 63, N. 7, P. 737 - 745, 2010.
21. Monteiro, G. T. R. Pesquisa em Saúde Pública: como desenvolver e validar instrumentos de coleta de dados. Curitiba: Appis, 2014.

22. Murphy PA. Sensitivity, Specificity, and Predictive Value Concepts for Clinicians. *Journal of Nurse-Midwifery*, v.39, 47–50, 1994.
23. NATIONAL RESEARCH COUNCIL & INSTITUTE OF MEDICINE [NRC & IM] Musculoskeletal disorders and the workplace: low back and upper extremities. Panel on musculoskeletal disorders and the workplace. Commission on behavioral and social sciences and education. Washington, DC: National Academy Press, 2001.
24. Namnik N, H Negahban, R Salehi, R Shafizadeh, MS Tabib. Validity and reliability of Persian version of the Specific Nordic questionnaire in Iranian industrial workers. *Work*, v.54, 35-41, 2016.
25. Ohlsson K, Attewell RG, Alm A, Johnsson B, Skerfving S. An assessment of neck and upper extremity disorders by questionnaire and clinical examination. *Ergonomics*, v.37, 891-897, 1994.
26. Punnett L and Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *J Electromyography Kinesiol.* 14:13-23, 2004.
27. Palmer K, Smith G, Kellingray S, Cooper C. Repeatability and validity of an upper limb and neck discomfort questionnaire: the utility of the standardized Nordic questionnaire. *Occup. Med.*, v. 49, n. 3, p. 171-5. 1999.
28. Pinheiro FA, Tróccolia BT, Carvalho CV. Validação do Questionário Nórdico de Sintomas Osteomusculares como medida de morbidade. *Revista de Saúde Pública*, São Paulo, v. 36, n. 3, p. 307-12. 2002.
29. Perreault N, Brisson C, Dionne CE, Montreuil S, Punnet L. Agreement between a self-administered questionnaire on musculoskeletal disorders of the neck-shoulder region and a physical examination. *BMC Musculoskeletal Disorders*, v. 9, n. 34. 2008.
30. Stal M, Moritz U, Johnsson B, Pinzke S . The natural course of musculoskeletal symptoms and clinical findings in upper extremities of female milkers. *Int J Occup Environ Health*, v.3, 190-197, 1997.
31. Silverstein BA, Stetson DS, Keyserling WM, Fine LJ. Work-related musculoskeletal disorders: comparison of data sources for surveillance. *Am J Ind Med*, v. 31, p. 600–8. 1997.
32. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable?. *Injury*, v. 42, 236-240, 2011.
33. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW.

Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, v. 19, n. 4, p. 539-549, 2012.

34. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, v. 60, n. 1, p. 34-42, 2007.
35. Takekawa KS, Gonçalves JS, Moriguchi CS, Coury HJCG, Sato TO. Can a self-administered questionnaire identify workers with chronic or recurring low back pain? *Industrial Health*, v. 53, 340 - 345. 2015.
36. Zheltoukhova K. Musculoskeletal Disorders and Work. Results of a survey of individuals living with Musculoskeletal Disorders in six European countries. 2013. Disponível em: <http://www.fitforworkeurope.eu/default.aspx.locid-0afnew009.htm>. Acesso em: 12 de fevereiro de 2017

VIII. ANEXOS

Anexo A- Versão ampliada do Questionário de Sintomas Musculoesqueléticos

Questionário Nórdico de Sintomas Musculoesqueléticos

Você teve dor ou desconforto ("dormência, formigamento, enrijecimento ou inchaço") em braços, mãos, pernas, pescoço ou região lombar durante os últimos doze meses? 1. () não 2. () sim
Se você respondeu SIM, por favor, complete a coluna para cada parte do corpo na qual surgiu a dor, nos dois quadros a seguir. Atenção: cada coluna diz respeito a uma parte do corpo descrita na primeira linha.

	Pescoço	Ombro	Cotovelo	Antebraço	Punho/mão	Parte alta das costas	Região lombar	Coxa	Joelho	Perna	Tornozelo	Pé
1. Que lado incomoda você? 1 – Direito 2 – Esquerdo 3 – Os dois												
2. Em que ano você notou o problema?												
3. Quanto tempo o problema dura geralmente? 1 - < de 1 hora 2 - > 1 hora até 1 dia inteiro 3 - >1 dia até 1 semana 4 - > 1 semana até 1 mês 5 - > 1 mês até 6 meses 6 - > 6 meses												
4. Quantos episódios do problema você teve? 1 – É constante, o tempo todo 2 – Diariamente 3 – Uma vez por semana 4 – Uma vez por mês 5 – A cada 2 ou 3 meses 6 – A cada 6 meses												
5. Você teve o problema nos últimos 7 dias? 1 – Sim 2 – Não												
6. Em uma escala de 0 a 5, como você classificaria o seu desconforto? Nenhum (0) \longrightarrow Insuportável (5)												
7. Você recebeu tratamento médico para o problema? 1 – Sim 2 – Não												
8. Quantos dias de trabalho você perdeu pelo problema?												
9. Quantos dias você ficou em trabalho leve ou restrito por causa do problema?												
10. Você mudou de trabalho por causa deste problema? 1 – Sim 2 – Não												
11. Você havia sofrido trauma agudo neste local (pancada, estirão, entorse, luxação)? 1 – Sim 2 – Não												

Traduzido por Fernandes, 2011.

Anexo B- Checklist COSMIN



The COSMIN checklist

Contact

CB Terwee, PhD
 VU University Medical Center
 Department of Epidemiology and Biostatistics
 EMGO Institute for Health and Care Research
 1081 BT Amsterdam
 The Netherlands
 Website: www.cosmin.nl, www.emgo.nl
 E-mail: cb.terwee@vumc.nl

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I
	Interpretability	Box J

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		yes	no	?
1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	<input type="checkbox"/>	<input type="checkbox"/>	
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency		yes	no	?
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Design requirements</i>		yes	no	?
2	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
4	Was the sample size included in the internal consistency analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	<input type="checkbox"/>	<input type="checkbox"/>	
6	Was the sample size included in the unidimensionality analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
9	for Classical Test Theory (CTT): Was Cronbach's alpha calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	for dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	for IRT: Was a goodness of fit statistic at a global level calculated? e.g. χ^2 , reliability coefficient of estimated latent trait value (index of (subject or item) separation)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)		yes	no	?
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were at least two measurements available?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Were the administrations independent?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Were patients stable in the interim period on the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8	Was the time interval appropriate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA ?
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	for ordinal scores: Was a weighted kappa calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box C. Measurement error: absolute measures				
	<i>Design requirements</i>	yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were at least two measurements available?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Were the administrations independent?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Were patients stable in the interim period on the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Was the time interval appropriate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	?
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	<input type="checkbox"/>	<input type="checkbox"/>	

Box D. Content validity (including face validity)				
<i>General requirements</i>		yes	no	?
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	

Box E. Structural validity				
		yes	no	?
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Design requirements</i>		yes	no	?
2	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
4	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
6	for CTT: Was exploratory or confirmatory factor analysis performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box F. Hypotheses testing				
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		yes	no	NA
5	Was the expected <i>direction</i> of correlations or mean differences included in the hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Was the expected absolute or relative <i>magnitude</i> of correlations or mean differences included in the hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	<input type="checkbox"/>	<input type="checkbox"/>	
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
10	Were design and statistical methods adequate for the hypotheses to be tested?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box G. Cross-cultural validity				
<i>Design requirements</i>		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	<input type="checkbox"/>	<input type="checkbox"/>	
6	Did the translators work independently from each other?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	Were items translated forward and backward?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Was there an adequate description of how differences between the original and translated versions were resolved?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Was the translation reviewed by a committee (e.g. original developers)?	<input type="checkbox"/>	<input type="checkbox"/>	
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	<input type="checkbox"/>	<input type="checkbox"/>	

11	Was the sample used in the pre-test adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the samples similar for all characteristics except language and/or cultural background?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
14	for CTT: Was confirmatory factor analysis performed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	for IRT: Was differential item function (DIF) between language groups assessed?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box H. Criterion validity				
	<i>Design requirements</i>	yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Can the criterion used or employed be considered as a reasonable 'gold standard'?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
	<i>Statistical methods</i>	yes	no	NA
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	for dichotomous scores: Were sensitivity and specificity determined?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box I. Responsiveness				
	<i>Design requirements</i>	yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was a longitudinal design with at least two measurement used?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	

7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	<input type="checkbox"/>	<input type="checkbox"/>	
Design requirements for hypotheses testing		yes	no	?
For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		yes	no	NA
9	Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Was an adequate description provided of the comparator instrument(s)?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the measurement properties of the comparator instrument(s) adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
14	Were design and statistical methods adequate for the hypotheses to be tested?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Design requirement for comparison to a gold standard		yes	no	?
For constructs for which a gold standard was available:				
15	Can the criterion for change be considered as a reasonable gold standard?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
<i>Statistical methods</i>		yes	no	NA
17	for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Box J. Interpretability		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	

3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was the distribution of the (total) scores in the study sample described?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the percentage of the respondents who had the lowest possible (total) score described?	<input type="checkbox"/>	<input type="checkbox"/>	
6	Was the percentage of the respondents who had the highest possible (total) score described?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Were scores and change scores (i.e. means and SD) presented for relevant (sub) groups? e.g. for normative groups, subgroups of patients, or the general population	<input type="checkbox"/>	<input type="checkbox"/>	
8	Was the minimal important change (MIC) or the minimal important difference (MID) determined?	<input type="checkbox"/>	<input type="checkbox"/>	
9	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	

Step 4: Determining the Generalisability of the results

Box Generalisability		yes	no	NA
	Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of:			
1	median or mean age (with standard deviation or range)?	<input type="checkbox"/>	<input type="checkbox"/>	
2	distribution of sex?	<input type="checkbox"/>	<input type="checkbox"/>	
3	important disease characteristics (e.g. severity, status, duration) and description of treatment?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	setting(s) in which the study was conducted? e.g. general population, primary care or hospital/rehabilitation care	<input type="checkbox"/>	<input type="checkbox"/>	
5	countries in which the study was conducted?	<input type="checkbox"/>	<input type="checkbox"/>	
6	language in which the HR-PRO instrument was evaluated?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Was the method used to select patients adequately described? e.g. convenience, consecutive, or random	<input type="checkbox"/>	<input type="checkbox"/>	
		yes	no	?
8	Was the percentage of missing responses (response rate) acceptable?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Anexo C- Escala de classificação de quatro pontos para o resultado da avaliação do Checklist COSMIN.

Step 1. Evaluated measurement properties in the article

	Internal consistency	Box A
	Reliability	Box B
	Measurement error	Box C
	Content validity	Box D
	Structural validity	Box E
	Hypotheses testing	Box F
	Cross-cultural validity	Box G
	Criterion validity	Box H
	Responsiveness	Box I

Step 2. Determining if the statistical method used in the article are based on CTT or IRT

Box General requirements for studies that applied Item Response Theory (IRT) models		excellent	good	fair	poor
		1	Was the IRT model used adequately described? e.g. One Parameter Logistic Model (OPLM), Partial Credit Model (PCM), Graded Response Model (GRM)	IRT model adequately described	IRT model not adequately described
2	Was the computer software package used adequately described? e.g. RUMM2020, WINSTEPS, OPLM, MULTILOG, PARSCALE, BILOG, NLMIXED	Software package adequately described	Software package not adequately described		
3	Was the method of estimation used adequately described? e.g. conditional maximum likelihood (CML), marginal maximum likelihood (MML)	Method of estimation adequately described	Method of estimation not adequately described		
4	Were the assumptions for estimating parameters of the IRT model checked? e.g. unidimensionality, local independence, and item fit (e.g. differential item functioning (DIF))	assumptions of the IRT model checked	assumptions of the IRT model partly checked	assumptions of the IRT model not checked or unknown	

To obtain a total score for the methodological quality of studies that use IRT methods, the 'worse score counts' algorithm should be applied to the IRT box in combination with the box of the measurement property that was evaluated in the IRT study. For example, if IRT methods are used to study internal consistency and item 4 in the IRT box is scored fair, while the items in the internal consistency box (box A) are all scored as good or excellent, the methodological quality score for internal consistency will be fair. However, if any of the items in box A is scored poor, the methodological quality score for internal consistency will be poor.

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency		excellent	good	fair	poor
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model? <i>Design requirements</i>				
2	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4	Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (< 30)
5	Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis NOT performed and no reference to another study
6	Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥ 100	5* #items and ≥ 100 OR 6-7* #items but < 100	5* #items but < 100	$< 5^*$ #items

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately		Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>				
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated	Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, 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
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2 , reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated		Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)					
		excellent	good	fair	poor
<i>Design requirements</i>					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4	Were at least two measurements available?	At least two measurements			Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated	
7	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
8	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate

9	Were the test conditions similar for both 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
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
11	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 with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box C. Measurement error: absolute measures				
	excellent	good	fair	poor
<i>Design requirements</i>				
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described	
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49) Small sample size (<30)
4	Were at least two measurements available?	At least two measurements		Only one measurement
5	Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent measurements NOT independent
6	Was the time interval stated?	Time interval stated		Time interval NOT stated
7	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
8	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate Time interval NOT appropriate

9	Were the test conditions similar for both 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
10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
11	for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population

Box D. Content validity (including face validity)					
		excellent	good	fair	poor
<i>General requirements</i>					
1	Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of the construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured

2	Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥ 10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3	Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4	Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehensively reflect the construct to be measured
5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study

Box E. Structural validity		excellent	good	fair	poor
1	Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
<i>Design requirements</i>					
2	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4	Was the sample size included in the analysis adequate?	7* #items and ≥ 100	5* #items and ≥ 100 OR 5-7* #items but < 100	5* #items but < 100	$< 5^*$ #items
5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study (e.g. rotation method not described)	Other important methodological flaws in the design or execution of the study (e.g. inappropriate rotation method)

<i>Statistical methods</i>			
6	for CTT: Was exploratory or confirmatory factor analysis performed?	Exploratory or confirmatory factor analysis performed and type of factor analysis appropriate in view of existing information	No exploratory or confirmatory factor analysis performed
7	for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?	IRT test for determining (uni)dimensionality performed	IRT test for determining (uni)dimensionality NOT performed

Box F. Hypotheses testing					
<i>Design requirements</i>		excellent	good	fair	Poor
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100 per analysis)	Good sample size (50-99 per analysis)	Moderate sample size (30-49 per analysis)	Small sample size (< 30 per analysis)

4	Were hypotheses regarding correlations or mean differences formulated a priori (i.e. before data collection)?	Multiple hypotheses formulated a priori	Minimal number of hypotheses formulate a priori	Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected
5	Was the expected <i>direction</i> of correlations or mean differences included in the hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated		
6	Was the expected absolute or relative <i>magnitude</i> of correlations or mean differences included in the hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated		
7	for convergent validity: Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Adequate description of most of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
8	for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s)

9	Were there any important flaws in the design or methods of the study? <i>Statistical methods</i>	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study	
10	Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Box G. Cross-cultural validity					
<i>Design requirements</i>		excellent	good	fair	poor
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	

3	Was the sample size included in the analysis adequate?	CTT: 7* #items and ≥100 IRT: ≥200 per group	CTT: 5* #items and ≥100 OR 5-7* #items but <100 IRT: ≥200 in 1 group and 100-199 in 1 group	CTT: 5* #items but <100 IRT: 100-199 per group	CTT: <5* #items IRT: (<100 in 1 or both groups
4	Were both the original language in which the HR-PRO instrument was developed, and the language in which the HR-PRO instrument was translated described?	Both source language and target language described			Source language NOT known
5	Was the expertise of the people involved in the translation process adequately described? e.g. expertise in the disease(s) involved, expertise in the construct to be measured, expertise in both languages	Expertise of the translators described with respect to disease, construct, and language	Expertise of the translators with respect to disease or construct poor or not described	Expertise of the translators with respect to language not described	
6	Did the translators work independently from each other?	Translators worked independent	Assumable that the translators worked independent	Unclear whether translators worked independent	Translators worked NOT independent
7	Were items translated forward and backward?	Multiple forward and multiple backward translations	Multiple forward translations but one backward translation	One forward and one backward translation	Only a forward translation
8	Was there an adequate description of how differences between the original and translated versions were resolved?	Adequate description of how differences between translators were resolved	Poorly or NOT described how differences between translators were resolved		

9	Was the translation reviewed by a committee (e.g. original developers)?	Translation reviewed by a committee (involving other people than the translators, e.g. the original developers)	Translation NOT reviewed by (such) a committee		
10	Was the HR-PRO instrument pre-tested (e.g. cognitive interviews) to check interpretation, cultural relevance of the translation, and ease of comprehension?	Translated instrument pre-tested in the target population	Translated instrument pre-tested, but unclear if this was done in the target population	Translated instrument pre-tested, but NOT in the target population	Translated instrument NOT pre-tested
11	Was the sample used in the pre-test adequately described?	Sample used in the pre-test adequately described	Sample used in the pre-test NOT (adequately) described		
12	Were the samples similar for all characteristics except language and/or cultural background?	Shown that samples were similar for all characteristics except language /culture	Stated (but not shown) that samples were similar for all characteristics except language /culture	Unclear whether samples were similar for all characteristics except language /culture	Samples were NOT similar for all characteristics except language /culture
13	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study		Other important methodological flaws in the design or execution of the study

<i>Statistical methods</i>		
14 for CTT: Was confirmatory factor analysis performed?	Multiple-group confirmatory factor analysis performed	Multiple-group confirmatory factor analysis NOT performed
15 for IRT: Was differential item function (DIF) between language groups assessed?	DIF between language groups assessed	DIF between language groups NOT assessed

Box H. Criterion validity				
	excellent	good	fair	poor
<i>Design requirements</i>				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (< 30)
4 Can the criterion used or employed be considered as a reasonable 'gold standard'?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'

5	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>				
6	for continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated		Correlations or AUC NOT calculated
7	for dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated		Sensitivity and specificity NOT calculated

Box I. Responsiveness					
		excellent	good	fair	poor
<i>Design requirements</i>					
1	Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2	Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3	Was the sample size included in the analysis adequate?	Adequate sample size (≥ 100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (< 30)
4	Was a longitudinal design with at least two measurement used?	Longitudinal design used			No longitudinal design used
5	Was the time interval stated?	Time interval adequately described			Time interval NOT described

6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	Anything that occurred during the interim period (e.g. treatment) adequately described	Assumable what occurred during the interim period	Unclear or NOT described what occurred during the interim period
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	Part of the patients were changed (evidence provided)	NO evidence provided, but assumable that part of the patients were changed	Unclear if part of the patients were changed Patients were NOT changed
Design requirements for hypotheses testing				
For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	Hypotheses formulated a priori		Hypotheses vague or not formulated but possible to deduce what was expected Unclear what was expected
9	Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected direction of the correlations or differences stated	Expected direction of the correlations or differences NOT stated	
10	Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	Expected magnitude of the correlations or differences stated	Expected magnitude of the correlations or differences NOT stated	

11 Was an adequate description provided of the comparator instrument(s)?	Adequate description of the constructs measured by the comparator instrument(s)	Poor description of the constructs measured by the comparator instrument(s)	NO description of the constructs measured by the comparator instrument(s)
12 Were the measurement properties of the comparator instrument(s) adequately described?	Adequate measurement properties of the comparator instrument(s) in a population similar to the study population	Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population
13 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study	Other minor methodological flaws in the design or execution of the study (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>			
14 Were design and statistical methods adequate for the hypotheses to be tested?	Statistical methods applied appropriate	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate

Design requirement for comparison to a gold standard					
For constructs for which a gold standard was available:					
15	Can the criterion for change be considered as a reasonable gold standard?	Criterion used can be considered an adequate 'gold standard' (evidence provided)	No evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'	Unclear whether the criterion used can be considered an adequate 'gold standard'	Criterion used can NOT be considered an adequate 'gold standard'
16	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
<i>Statistical methods</i>					
17	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
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated

Interpretability

We recommend to use the Interpretability box to extract all information on the interpretability issues described in this box of the instruments under study from the included articles.

Box Interpretability	
Percentage of missing items	
Description of how missing items were handled	
Distribution of the (total) scores	
Percentage of the respondents who had the lowest possible (total) score	
Percentage of the respondents who had the highest possible (total) score	
Scores and change scores (i.e. means and SD) for relevant (sub) groups, e.g. for normative groups, subgroups of patients, or the general population	
Minimal Important Change (MIC) or Minimal Important Difference (MID)	

Generalizability

We recommend to use the Generalizability box to extract data on the characteristics of the study populations and sampling procedures of the included studies.

Box Generalisability	
Median or mean age (with standard deviation or range)	
Distribution of sex	
Important disease characteristics (e.g. severity, status, duration) and description of treatment	
Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)	
Countries in which the study was conducted	
Language in which the HR-PRO instrument was evaluated	
Method used to select patients (e.g. convenience, consecutive, or random)	
Percentage of missing responses (response rate)	

Anexo D- Protocolo da revisão sistemática

PROSPERO International prospective register of systematic reviews

Validity and reliability of the Nordic Musculoskeletal Questionnaire: a systematic review

Ivone Batista Alves, Verônica Lima

Citation

Ivone Batista Alves, Verônica Lima. Validity and reliability of the Nordic Musculoskeletal Questionnaire: a systematic review. PROSPERO 2016:CRD42016051213 Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42016051213

Review question(s)

What is the validity and reliability of the Nordic Musculoskeletal Questionnaire to assess work-related musculoskeletal disorders?

Searches

The research will be carried out using databases online (CINAHL, MEDLINE, Scopus, Web of Science) and will select articles published since September 1987 to January 2017.

The search will be restricted to studies published in English or Portuguese.

Types of study to be included

We will include validity studies (criterion validity) and/or reliability studies of the Nordic Musculoskeletal Questionnaire in worker population.

Condition or domain being studied

Musculoskeletal disorders in workers.

Participants/ population

Groups of workers.

Intervention(s), exposure(s)

None

Comparator(s)/ control

Not applicable.

Outcome(s)

Primary outcomes

To characterize the measurement properties of the Nordic Musculoskeletal Questionnaire (NMQ) and to evaluate the methodological quality of the studies;

Secondary outcomes

None

Data extraction, (selection and coding)

Through the title and/or abstracts of studies available from the search strategy. These will be screened by one review author in order to identify studies that potentially meet the inclusion criteria. The full text of these potentially eligible studies will be assessed for eligibility by two reviewers. Disagreements between the reviewers will be resolved by discussion.

Extracted information will include: characteristics of the studies; methodological quality; study completion rates.

Risk of bias (quality) assessment

Reading, assessment and data extraction will be carried out independently by two reviewers.

Strategy for data synthesis

Aggregate data will be extracted and a quantitative synthesis is planned.

Analysis of subgroups or subsets

Not applicable.

Dissemination plans

The research will be published in peer-reviewed literature

Contact details for further information

Mrs Batista Alves

Rua Santa Clara, 143, Ed: Santa Amelia, apt 402, Nazaré, Salvador-BA, CEP: 40040450. Brazil

vonyssa@bol.com.br

Organisational affiliation of the review

Ivone Batista Alves (Student); Dr. Veronica Cadena Maria Lima, Department of Statistics at the Federal University of Bahia, Brazil.

Review team

Mrs Ivone Batista Alves, Student at Federal University of Bahia

Dr Verônica Lima, Federal University of Bahia

Anticipated or actual start date

10 November 2016

Anticipated completion date

30 January 2017

Funding sources/sponsors

Coordination for the Improvement of Higher Education Personnel (CAPES)

Conflicts of interest

None known

Language

Portuguese-Brazil

Country

Brazil

Subject index terms status

Subject indexing assigned by CRD

Subject index terms

Humans; Musculoskeletal Diseases; Reproducibility of Results; Surveys and Questionnaires

Stage of review

Ongoing

Date of registration in PROSPERO

10 November 2016

Date of publication of this revision

10 November 2016

Stage of review at time of this submission	Started	Completed
Preliminary searches	No	No
Piloting of the study selection process	No	No
Formal screening of search results against eligibility criteria	No	No
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

PROSPERO

International prospective register of systematic reviews

The information in this record has been provided by the named contact for this review. CRD has accepted this information in good faith and registered the review in PROSPERO. CRD bears no responsibility or liability for the content of this registration record, any associated files or external websites.

IX- Apêndice

Estratégias de busca utilizada por base de dados.

Base	Estratégias utilizadas
Medline/Pubmed	<p>((validity) OR validation) AND Nordic Musculoskeletal Questionnaire</p> <p>((Sensibility) OR Specificity) AND Nordic Musculoskeletal Questionnaire</p> <p>((((Reproducibility) OR Repeatability) OR Reliability) AND Nordic Musculoskeletal Questionnaire)</p>
Web of science	<p>(validity) AND Tópico: (Nordic Musculoskeletal Questionnaire)</p> <p>(reliability) AND Tópico: (Nordic Musculoskeletal Questionnaire)</p> <p>(Repeatability) AND Tópico: (Nordic Musculoskeletal Questionnaire)</p> <p>Sensibility OR Specificity) AND Tópico: (Nordic Musculoskeletal Questionnaire)</p> <p>(Reproducibility) AND Tópico: (Nordic Musculoskeletal Questionnaire)</p>
Scopus	<p>TITLE-ABS-KEY (validity) AND TITLE-ABS-KEY (nordic musculoskeletal questionnaire) AND PUBYEAR > 1986</p> <p>TITLE-ABS-KEY (sensibility OR specificity) AND TITLE-ABS-KEY (nordic musculoskeletal questionnaire)</p> <p>(TITLE-ABS-KEY (reliability) OR TITLE-ABS-KEY (repeatability) OR TITLE-ABS-KEY (reproducibility) AND TITLE-ABS-KEY (nordic musculoskeletal)</p>
CINHL	<p>validity AND nordic musculoskeletal questionnaire</p> <p>Limitadores</p> <p>Remove Data de publicação: 19870901-20161131</p> <p>validation AND nordic musculoskeletal questionnaire</p> <p>reliability AND nordic musculoskeletal questionnaire</p>